On the Relationship of Server Disk Workloads and Client File Requests

John R. Heath*
Department of Computer Science
University of Southern Maine
Portland, Maine 04103

Stephen A.R. Houser
University Computing Technologies
University of Southern Maine
Portland, Maine 04103

In this study, we consider the relationship of client IO requests submitted to a file server and the server's disk subsystem IO traffic. We collected traces of client-server IOs and, during each trace period, we also traced server storage subsystem workloads. We analyze and compare trace pairs collected in each trace period. We evaluate file server performance and investigate the relationship between client file requests and server disk workloads.

1 Introduction

Network file servers provide global file systems that are shared by client workstations. To provide acceptable performance levels, servers employ a large file cache and, using a variety of algorithms, attempt to anticipate future client requests and store, in file cache, data that are likely to be accessed. In an earlier study [HEAT95], the authors analyzed server disk workload traces and characterized server disk subsystem performance. In this study, we traced client IO requests made to network file servers, and, during the same period, we also traced server disk workloads. We then analyzed and compared the traces for the purpose of acquiring a better understanding of the relationship between client IO requests and server storage system workloads. Understanding the relationship between client-server traffic and server disk workloads is useful in predicting server storage requirements and planning for future growth of server storage subsystems. Our results are also useful for parameterizing simulation and queueing network models.

In the next section, we describe the systems traced and our trace methodology. In section 3, we compare client IO request mix with the IO mix of server subsystem workloads. We then, in section 4, investigate the relationship between client-server throughput and server disk subsystem throughput. In section 5, we evaluate the effectiveness of the server file cache by analyzing client IO response times. In the last section we provide concluding remarks.

2 Trace Environment and Methodology

Disk workload traces analyzed in this study were collected in April 1994 over a period of several days from two file servers on different LANs. One server provided the file system for seventy diskless workstations connected to a general access LAN used primarily by university students from all academic disciplines. The most frequently used server application was word processing, accounting for 60% of all connect time. Other applications included spreadsheet programs, communications services, programming, MS-DOS, database, and courseware. The LAN is a hub-configured Ethernet with workstations interconnected by twisted-pair cable. The LAN's single file server runs Novell NetWare and uses two SCSI disk drives to support its file system.

The second set of traces was obtained from a network file server used by administrative staff in several university departments. Client workstations have local file systems and use the server for shared applications, primarily word processing, and access to shared data. During working hours, the number of workstations connected to the server varies from 50 to 75. The administration LAN is also hub-configured Ethernet and the file server has essentially the same configuration as the student file server.

Traces of client IO requests were collected by a network traffic monitor. The monitor, which ran on a dedicated workstation, collected transmissions to and from the server. Frame headers, containing the requested file operation, and the time each frame was received were stored in the trace file. Table 1 gives, for each trace, the total number of NetWare Core Protocol (NCP) operations observed, the number of NCP file operations observed, and the percentage of file operations to total NCP operations. In addition to file commands, the NCP count includes such non-file operations as server advertising, queue and system maintenance, and authentication services. The monitor was able to collect packets for about a two-hour period on the student network without exceeding its storage limits and for four hours on the more lightly loaded administration network.

* This work was supported in part by Digital Equipment Corporation, Shrewsbury, MA.

Trace	Time Period	Total NCP Operations	NCP File Operations	% File Operations
Student Access File Server				
MON	9:00 am -11:00 am	638331	602274	94.3%
WED	9:00 am -11:00 am	1959233	1882269	96.1%
FRI	9:00 am -11:00 am	1223197	1148940	93.9%
Totals		3820761	3633483	95.1%
Administration File Server				
MON	8:00 am -12:00 noon	925182	825197	89.2%
WED	8:00 am -12:00 noon	923957	813667	88.0%
THU	8:00 am -12:00 noon	702327	604689	86.1%
FRI	8:00 am -12:00 noon	619159	508264	82.1%
Totals		3170625	2751817	86.8%

Table 1: Network Trace Summary

Trace	Read%	Write%	Open%	Close%	Other%
Student Access File Server					
MON	80.4%	8.5%	3.3%	2.8%	5.0%
WED	78.7%	12.2%	2.6%	2.4%	4.1%
FRI	76.7%	10.8%	3.7%	3.4%	5.4%
Overall	78.4%	11.2%	3.1%	2.8%	4.6%
Administration File Server					
MON	70.5%	3.4%	3.7%	3.4%	19.1%
WED	68.8%	4.8%	5.9%	4.4%	16.1%
THU	70.6%	4.9%	6.8%	4.8%	13.0%
FRI	65.4%	5.1%	5.3%	4.8%	19.3%
Overall	69.1%	4.5%	5.3%	4.2%	16.9%

Table 2: NCP File Operation Statistics

A SCSI bus monitor [PEER93], installed in an MS-DOS PC, was used to trace server disk storage workloads. The monitor attached directly to the server's SCSI bus and recorded all commands transmitted on the bus. Traces were collected during what were generally believed to be the busiest periods, from 9:00am to 11:00am on the student network and from 8:00am to 12:00n on the administration network. Trace data were interpreted and analyzed after the tracing was completed. Both monitors were passive and did not affect system performance.

3 IO Request Mix

In this section, we consider the mix of file operations submitted by network clients to the servers and compare the mix with that of IO requests submitted by the servers to their disk storage subsystems. Table 2 shows the percentage of each NCP file operation type sent to the two servers. In both networks, the great majority of file requests were read operations, 78% in the student network, 69% in the administration network. We note that the mean number of IOs for each file open is 30 in the student network and only half that in the administration network. A possible explanation for this difference is that workstations in the student network are diskless, and rely on the file server for all file operations, whereas the administration network workstations have local disks that can be used for paging, storing temporary files and system binaries.

File write operations are a small percentage of all file requests, 11% in the student network and only 4.5% in the administration network. NCP file operations other than read, write, open, and close are classified as *Other* in Table 2. These include NCP operations such as Create File, Erase File, Rename File, Get File Size, File Search, Set Attributes, to name a few. We note that in contrast to the student network, the percentage of *Other* file operations on the administration network is rather high, 17%. The cause of this difference is not clear; possibly it is because workstations with local disks handle many reads and writes locally, once a file is opened and initially read from the server. Therefore they exhibit a lower percentage of read and write NCP operations.

If we consider read and write operations only, ignoring all other file operations, the mean read-to-write ratio is 7:1 for the student network and 16:1 for the administration network. However, NCP file operations cate-

Trace	Client Request R/W Ratio	Server Disk R/W Ratio
Student File Server		
MON	7.63	0.33
WED	5.73	0.31
FRI	6.89	0.24
Mean	6.75	0.29
Administration File Server		
MON	13.25	1.07
WED	9.65	1.08
THU	9.07	0.87
FRI	8.78	0.81
Mean	10.19	0.96

Table 3: Client and Server Disk Read/Write Ratios

gorized as *Other* also require file IOs to execute. If we include read and write operations required by all NCP operations, the mean read-to-write ratio for the student network remains about 7:1, but is reduced to 10:1 for the administration network. In Table 3, we compare the read-to-write ratio of client file operations with the read-to-write ratio of server disk subsystem IOs. The significant difference in these two ratios demonstrates the servers' effectiveness in processing client read requests without accessing the disk subsystem. In the student network, the read-to-write ratio of client file requests is about 7:1, but the read-to-write ratio of the server disk subsystem trace is 1:3. In the administration network, client file requests have a 10:1 read-to-write ratio, whereas the disk subsystem workload has a 1:1 read-to-write ratio.

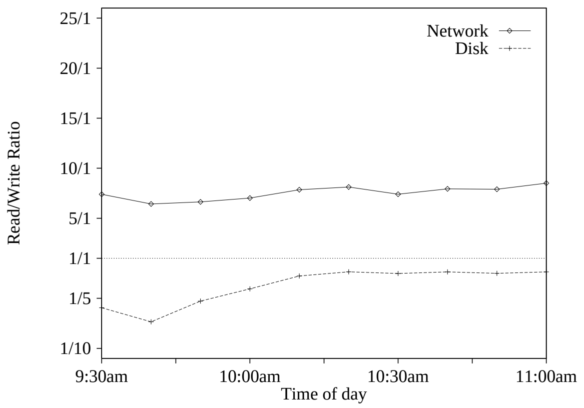

Figure 1: Student File Server Read/Write Ratio, MON

Read-to-write ratios for the past 30 minutes, computed at 10 minute intervals, are plotted in Figs. 1-7. Each figure shows read-to-write ratios for both client file requests and for server disk IOs during a trace period. Figs. 1-3 show, for the three trace periods, read-to-write ratios for student network. We observe that network and disk read-to-write ratios display little fluctua-

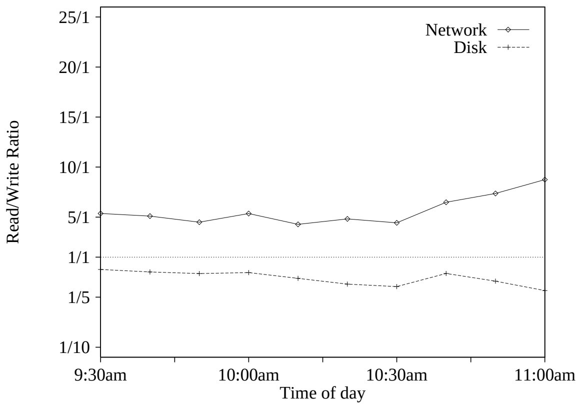

Figure 2: Student File Server Read/Write Ratio, WED

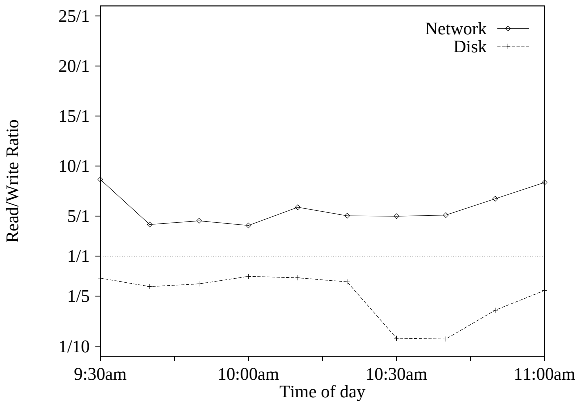

Figure 3: Student File Server Read/Write Ratio, FRI

tion throughout the trace periods. In all three traces, client file request ratios remain between 5:1 and 10:1, while read-to-write ratios of server disk traces remain generally between 1:2 and 1:5, the exception being the end of the Friday trace which has a read-to-write ratio of 1:9, Fig. 3.

Graphs shown in Figs. 4-7 depict read-to-write ratios for administration network traces. Although there is much fluctuation in client trace read-to-write ratios, ranging from 5:1 to 20:1, disk workload read-to-write ratios remain relatively uniform at about 1:1. Examination of the ratios indicates there is not a strong correlation between server disk subsystem read-to-write ratios and client file request ratios. Furthermore, although client file requests are predominantly read requests, server disk workloads consists of as many or more writes than reads. Clearly, server disk workloads have a very different IO mix than local file systems workloads reported in the literature [BISW90, OUST85, RAMA92, SMIT85] which have ratios more similar to those observed in the client-server traces. Our results also indicate that the file servers, through their file caches and read-ahead algorithms, service a large percentage of read operations directly from file cache without accessing the disk subsystem.

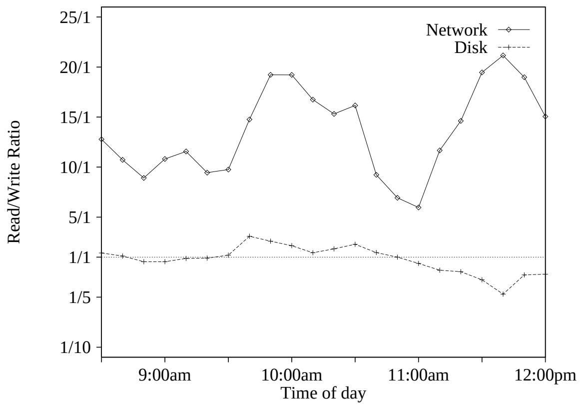

Figure 4: Administration File Server Read/Write Ratio, MON

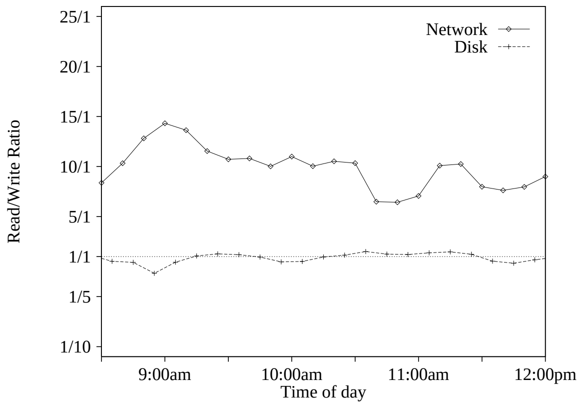

Figure 5: Administration File Server Read/Write Ratio, WED

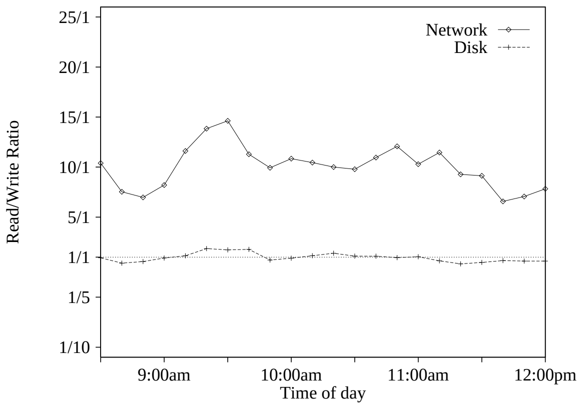

Figure 6: Administration File Server Read/Write Ratio, THU

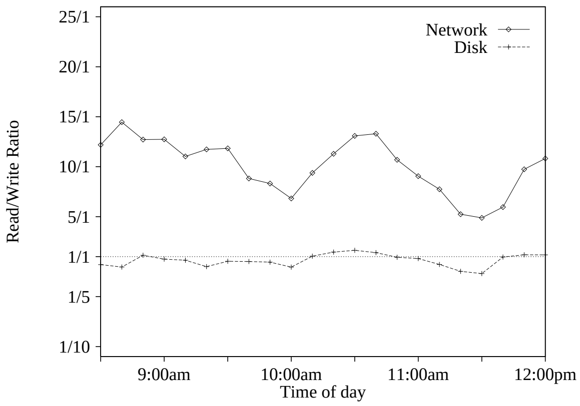

Figure 7: Administration File Server Read/Write Ratio, FRI

4 Throughput

In this section, we compare client IO throughput (measured IO rate) with server disk subsystem throughput during the same trace period. In our throughput analysis, the output unit is IOs. The reader should be aware that the amount of data transferred per IO is different for the two streams, client IO and server disk IO. Client file IOs are limited in length by the underlying network protocols. Specifically, the network packet payload limits IOs to no more than one kilobyte, regardless of the length of the client application's original request. Furthermore, any request that is routed through the internetwork router is segmented into requests of no more than 512 bytes. Consequently, the distribution of client IO request lengths, particularly reads, have a high percentage of 1024 and 512-byte sizes; see Figs. 8 and 9. Write IO lengths, in addition to 512 and 1024 byte sizes, which are common, also include many smaller lengths; on the administration network, client write request lengths of 64, 128, and 192 bytes account for about 15% each of all client writes. See Figs. 10 and 11.

Servers read ahead of the requested data in anticipation of future client read requests. Consequently, the length of read requests submitted by a server to its storage system are larger than the client's request. Novell NetWare, the network operating system that runs on the servers studied, issues disk read requests that are multiples of 4 kilobytes, regardless of the size of the client request. In fact, all file reads submitted by the administration server were for 4 kilobytes. In the student network, ninety percent of the reads submitted by the server were for 12 kilobytes and all other read requests were for either 4 or 8 kilobytes. Also, disk write request lengths are for at least 512 bytes, the disk block size, and, in fact, three-fourths of write operations submitted by both servers were for 512 bytes. In the remainder of this section, we compare client-server IO throughput with server disk subsystem throughput.

Table 4 lists throughput, during each trace period, of both network interface and server subsystem. The rows labeled *mean* list the average trace throughputs for the two systems. From the Table 4, we see that, on average, on the student network, 152 client reads/sec resulted in only 2.5 disk reads/sec and that 24 client writes/sec resulted in 8 disk writes/sec. In the adminis-

Trace	Read		Write		Overall	
	disk IOs/s	NCP IOs/s	disk IOs/s	NCP IOs/s	disk IOs/s	NCP IOs/s
Student File Server						
MON	2.45	74.04	7.50	9.71	9.95	83.75
WED	3.64	222.83	10.67	38.88	14.31	261.71
FRI	1.44	159.79	5.95	23.20	7.39	159.79
Mean	2.51	152.22	8.04	23.93	10.55	168.42
Administration File Server						
MON	1.07	38.15	1.63	2.88	3.38	41.03
WED	2.07	37.28	1.92	3.86	3.99	41.13
THU	1.28	28.85	1.47	3.18	2.75	32.03
FRI	1.07	21.16	1.32	2.41	2.40	23.56
Mean	1.54	31.36	1.59	3.08	3.13	34.44

Table 4: Client and Disk Throughput

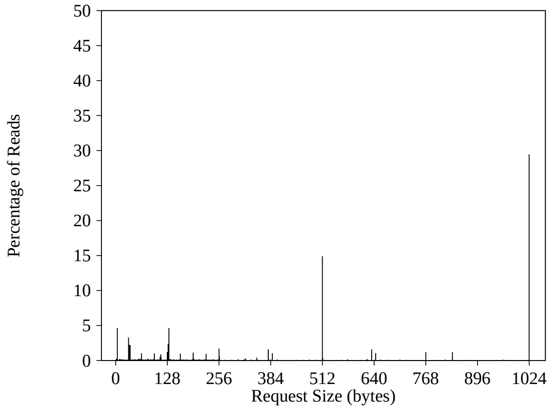

Figure 8: NCP Read Size Distribution, Student File Server

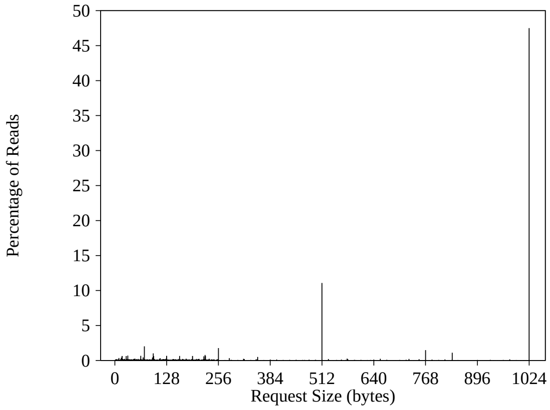

Figure 9: NCP Read Size Distribution, Administration File Server

tration network, we observe that 31 client reads/sec resulted in disk throughput of 1.5 disk reads/sec; 3 client writes/sec resulted in 1.5 disk writes/sec. Clearly, the server file cache significantly reduces disk reads and, although to a lesser degree, also reduces disk writes.

Measured throughputs for 30 minute periods, computed at 10 minute intervals, are graphed in Figs. 12-18. Each figure has two curves, one shows client-server IO throughput, the other shows server disk subsystem IO throughput. There is one figure for each trace period. Note, the vertical axis (throughput) scale is logarithmic.

We observe in all figures that network interface throughput is an order of magnitude greater than server disk subsystem throughput. Examination of the figures suggests a strong correlation between client file IO and server subsystem IO. To quantify these apparent correlations, we computed correlation coefficients of network interface throughputs and disk subsystem throughputs for each trace period. The results are given in Table 5, column 2. We also computed correlation coefficients for read and write throughputs, columns 3 and 4. All traces, with the exception of the MON trace on the administration server, show a strong correlation between client-server IO and server disk IO with write output being more strongly correlated than read output.

5 Response Time

To be effective, a file server must respond to client requests in times comparable to that of local file systems. Recall, adequate response time levels are achieved by storing recently accessed files or parts of recently accessed files in a region of the server's memory, called the file cache. Read IO requests for records stored in the cache are serviced directly from the file cache without accessing the server's disk storage, thereby substantially reducing response time. Disk write requests are also stored in the file cache, and the actual write to disk is delayed. The purpose is to reduce disk writes by processing multiple writes to cached data without writing to disk, and by collecting writes to consecutive blocks into a single disk write.

In this section, we present distributions of response time measurements of client requests and use these measurements to assess the effectiveness of the

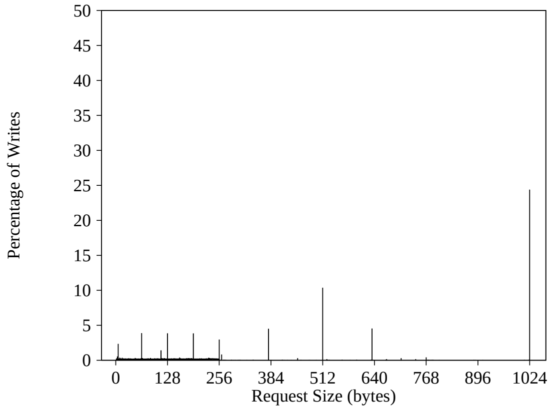

Figure 10: NCP Write Size Distribution, Student File Server

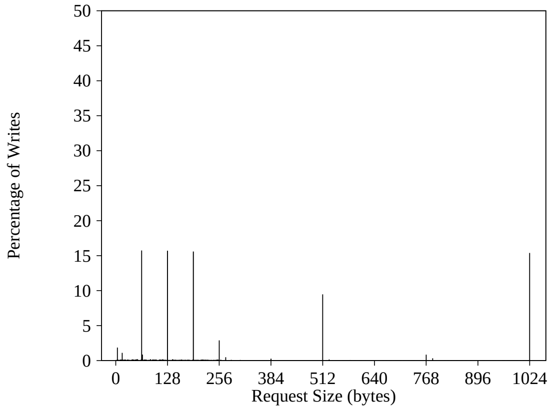

Figure 11: NCP Write Size Distribution, Administration File Server

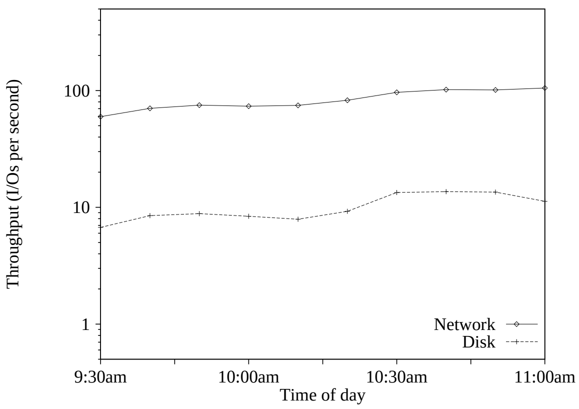

Figure 12: Student File Server Throughput, MON

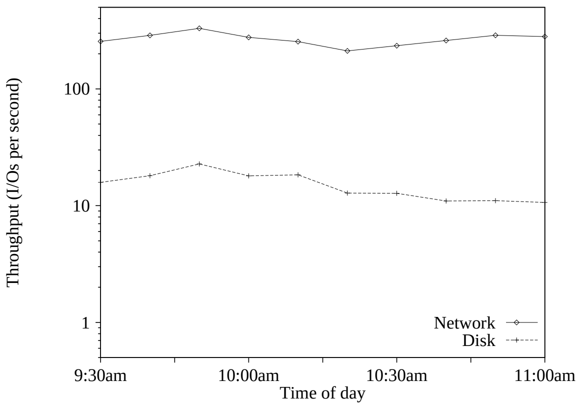

Figure 13: Student File Server Throughput, WED

servers' file caches in reducing response times. Response times of client IOs were measured in the following manner. The network monitor recorded the time each frame was received. From these time stamps, we computed time from the start of each operation until the start of the server reply packet, as seen by the monitor. In the discussion that follows, we refer to these times as response times. By examining the distribution of these times, we can estimate the percentage of file cache hits and assess the effectiveness of the server in quickly processing client IOs.

We first consider read response times. Client read requests may be handled in one of several ways. If the requested data are stored in the server's file cache, the server can send the data immediately without disk access. If the data are not found in the file cache, the disk subsystem is accessed. The data may be in the disk controller cache, or if it is not, the disk is accessed and the data read. Measurements of the server disk subsystems' response times , reported in [HEAT95], show the disk subsystem's minimum response time is between 3 and 4ms, a response time that results when the requested data are stored in the disk controller cache. Consequently, we can assume that client read requests with response time less than 3ms were serviced from the server file cache. In the student network, 97% of client read requests had response times less than 3ms. In the administration network, 95% of client read requests had response times less than 3ms. Histograms of read response times under 5ms for the student and administration networks are shown in Figs. 19 and 20, respectively. Each bar represents a 0.1ms interval. For example, the bar labeled 0.5 represents the percentage of requests with response times between 0.5 and $5.9\bar{9}$ms. Read request response times for all traces, of the specified network, are included in the histograms. We estimate that with the additional delays imposed by transmission times, server processing time, and queueing time included, response times under 5ms result from a file cache hit. The cache hit percentage for each network increases by one percent when we include response times between 3 and 5ms.

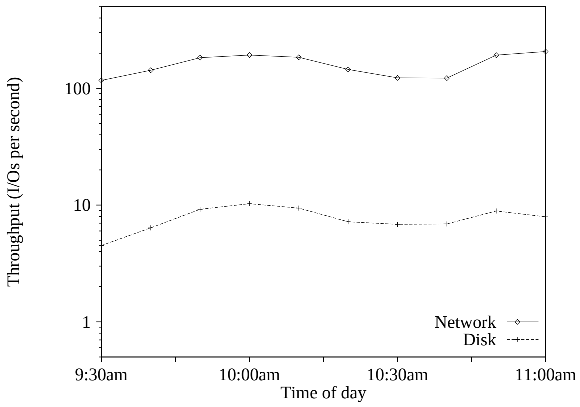

Figure 14: Student File Server Throughput, FRI

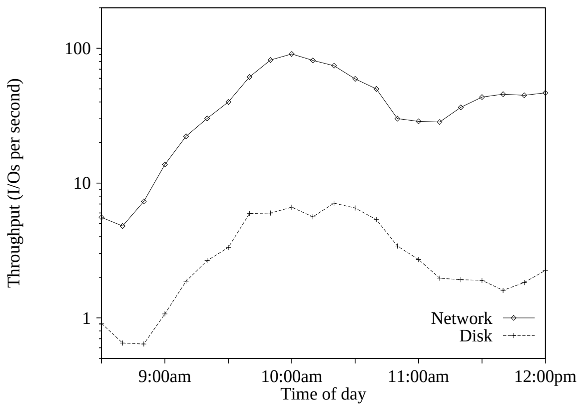

Figure 15: Administration File Server Throughput, MON

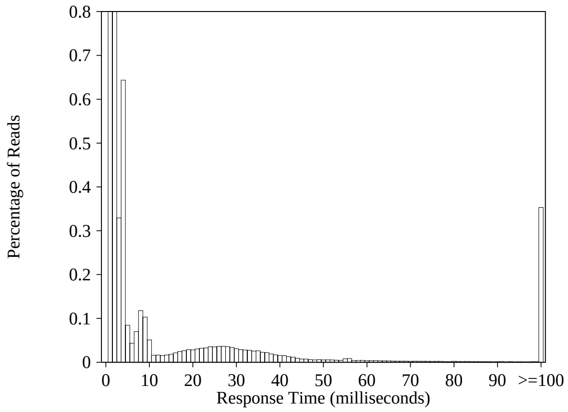

Figure 21: Network Read Request Response Time Histogram, Student File Server

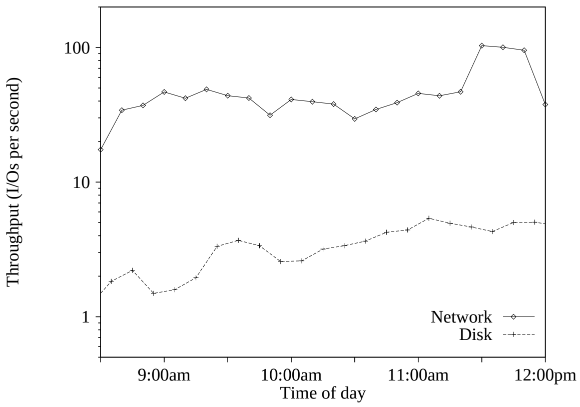

Figure 16: Administration File Server Throughput, WED

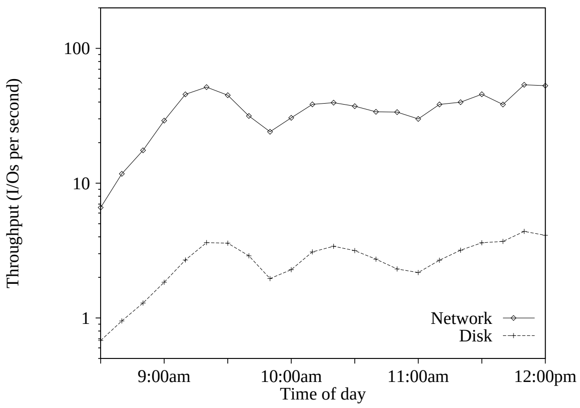

Figure 17: Administration File Server Throughput, THU

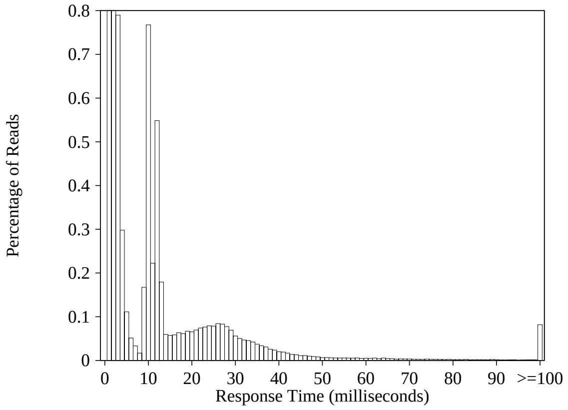

Figure 22: Network Read Request Response Time Histogram, Administration File Server

A histogram of student network client read response times up to 100ms is shown in Fig. 21. Each bar represents a 1ms interval. Response times greater than or equal to 100ms are collected in a single bar. The histogram is intended to emphasize the distribution of response times greater than 5ms; hence, the vertical axis is scaled in such a way that many per-

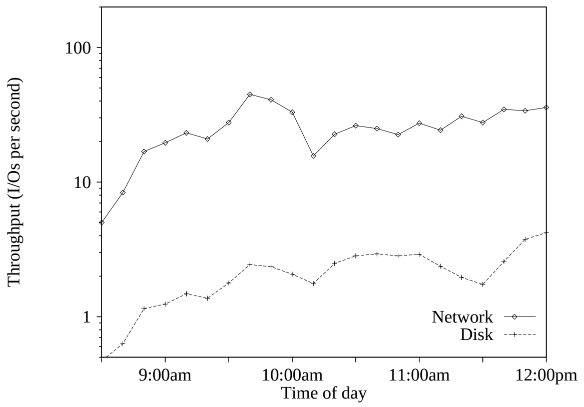

Figure 18: Administration File Server Throughput, FRI

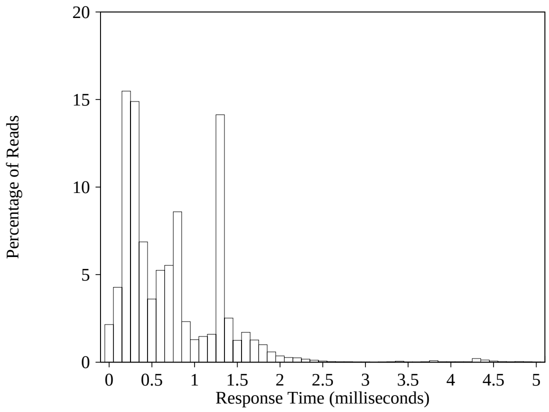

Figure 19: Network Read Request Response Time Histogram, <=5ms, Student File Server

Correlation Coefficients			
Trace	Overall	Read	Write
Student File Server			
MON	0.964	0.898	0.975
WED	0.812	0.614	0.977
FRI	0.945	0.754	0.902
Administration File Server			
MON	0.883	0.798	0.924
WED	0.552	0.342	0.924
THU	0.957	0.898	0.762
FRI	0.776	0.682	0.710

Table 5: Correlation Coefficients of Client and Disk Subsystem Throughput

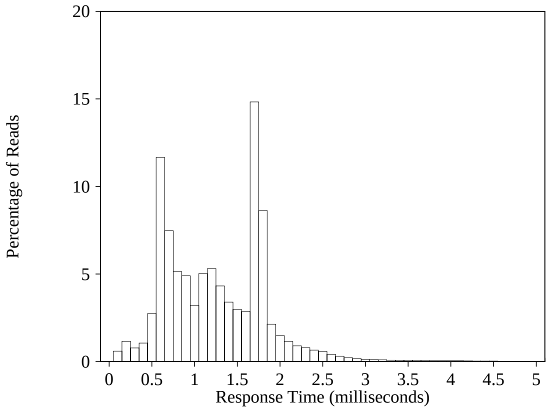

Figure 20: Network Read Request Response Time Histogram, <=5ms, Administration File Server

centages associated with response times less than 5ms are not shown in the figure. In Fig. 21, we observe a secondary response time peak between 7 and 10ms intervals. Based on service time measurements of similar disks [HEAT93], as well as measured disk response times [HEAT95], we attribute these response times to accesses serviced by controller cache or disk accesses that do not require a seek. This region contains 0.3% of all requests. We observe an even more pronounced secondary peak in the administration network histogram, Fig. 22. The percentage of read response times contained in this region, which we attribute to controller cache accesses and disk accesses requiring no seek or, possibly, a very short seek, is 2% of client reads. The percentage of requests requiring a disk seek appears to be only 1.3% in the student network and 1.9% in the administration network.

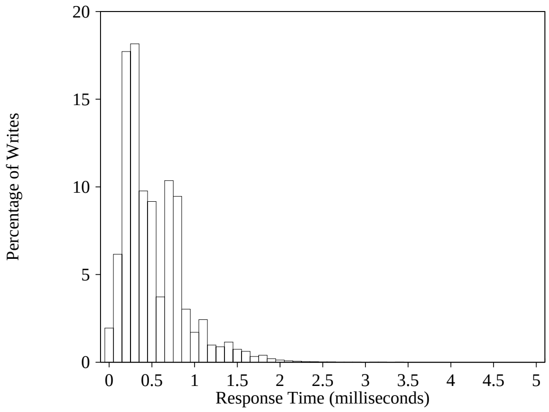

Figure 23: Network Write Request Response Time Histogram, Student File Server

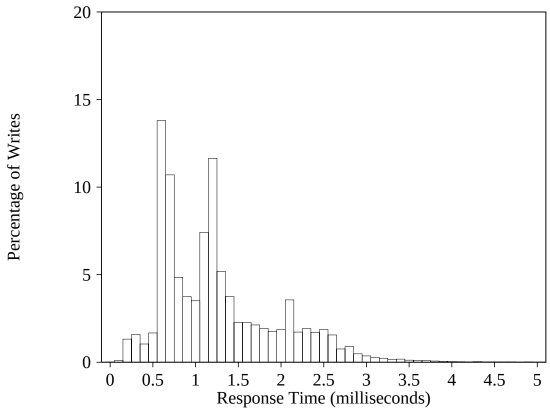

Figure 24: Network Write Request Response Time Histogram, Administration File Server

Response time histograms for client write requests are shown in Figs. 23 and 24. The server employs a *write-back* policy, referred to above, in which write data is delayed in the server cache before being written to disk. The server sends a response acknowledgment to the source client when the write data is cached, before it is written to disk. Therefore, write response times do not include disk write delays and, consequently, as we observe in the figures, nearly all write request response times are relatively short. In the student network, 99.4% of write response times are less than 2ms; in the administration network, 98.7% are less than 3ms.

6 Concluding Remarks

We have analyzed and compared client IO workload traces and server disk workload traces. We observed that client IOs are predominantly file reads, while server disk workloads had an equal or greater number of writes than reads. There appeared to be no clear correlation between client traffic read-to-write ratios and disk subsystem read-to-write ratios. We found client IO throughput to be an order of magnitude greater than server disk throughput. Client request response times indicated that nearly all reads are serviced by the server from its file cache.

Further study of additional client IO and server disk traces is needed to determine whether or not network trace data can be used to accurately predict server disk subsystem workloads. However, our analysis suggests that only a small percentage of client reads, less than 5%, require disk access. Also, there is, generally, a strong correlation between client IO demand and server disk throughput.

References

[BISW90] P. Biswas and K.K. Ramakrishnan. File Characterizations of VAX/VMS Enviroments. In *Proc. 10th International Conference on Distributed Computing Systems*, pages 227–234, May 1990.

[HEAT93] John R. Heath "Measurement and Performance Evaluation of Seagate's Elite 3 Disk Drive" Technical Report TR 93-5, Dept. of Computer Science, University of Southern Maine, May 1993.

[HEAT95] John R. Heath and Stephen A. R. Houser. "Analysis of Disk Workloads in Network File Server Environments" *Proc. CMG95*, pages 313–322, Dec. 1995.

[OUST85] John K. Ousterhout and et.al. A trace driven analysis of the UNIX 4.2 BSD file system. *Proc. Tenth Sympos. on Op. Sys. Princ.*, pages 15–24, Dec. 1985.

[PEER93] Peer Protocols Inc. *Peer Protocol SCSI Analyzer*, 1993.

[RAMA92] K.K. Ramakrishnan, P. Biswas, and R. Karedla. Analysis of File I/O Traces in Commercial Computing Enviroments. In *Proc. 1992 ACM Sigmetrics & Performance*, pages 78–90, June 1992.

[SMIT85] Alan J. Smith. Disk-cache – miss ratio analysis and design considerations. *ACM Trans. on Comp. Sys.*, pages 161–203, August 1985.