March 25, 1985

A BENCHMARK COMPARISON OF SEVERAL DIFFERENT SIZED COMPUTERS, ALL USING EITHER UNIX OR ANOTHER SIMILAR OPERATING SYSTEM

Prepared by:

George Robertson

Introduction

A magazine article published by BYTE magazine, August 1984, was used as the basis for this report. The BYTE article included data for several popular "multi-user UNIX" systems. The original data from the article was duplicated in this paper. The BYTE data included a TRS-80 model 16b, 384K with an internal 15MB Hard Disk. Tandy's Computer Product Engineering group added the next set of systems to the benchmark. These were the two Convergent Technologies "Mini-Frame" systems along with a TRS-80 Model 16b, 512K with an internal 15MB Hard Disk. The CT Mini Frames had demand paged virtual memory enhancements on Unisoft's UNIX System 5 running on a Motorola 68010.

For this report, the first effort was just to test the AT&T 3B2/300 system to try to find whether it might have some future usefulness at Tandy Corporation. After running the original BYTE benchmark programs as part of that evaluation, it was then decided to compare the AT&T system with the data from BYTE. In adding the information for the AT&T system, it also became reasonable to add similar test data for the newer machines now available. These new machines are the Tandy 2000, the IBM PC AT, and the Tandy 6000, all with a version of Xenix 3.0. One older TRS-80 model 16 "A" (smaller case) was "dug up" so that it could be used as an appropriate way to look at any performance improvements that have taken place over the last few years. The third model 16b was added to show the effects of larger real memory, if any.

The information for the machines in the last section of the table below was done in the hope of getting good comparison data for our entire Xenix product family. This report is also one of the first comprehensive tests to include the <u>Motorola 68000</u> & <u>68010</u> as well as the <u>Intel 8088</u>, <u>80186</u>, and <u>80286</u>.

	System	Version	RAM size	Disk	CPU	MHz	Retail Price
#1	VAX 11/780	4.1 BSD	4096K	2-256 MB	VAX	?	~$350,000
	Masscomp	UNIX Sys III+	2048K	1- 50 MB	68010	10	~$ 30,000
	Sun 2/120	4.2 BSD	2048K	1- 42 MB	68010	8	~$ 25,000
	VAX 11/750	4.1 BSD	2048K	1-121 MB	VAX	?	~$185,000
	PDP 11/70	2.8 BSD	1536K	1-400 MB	DEC	?	~$180,000
	Altos 986	Xenix	1024K	1- 40 MB	68000	6	~$ 12,000
	IBM XT PC/IX	PC/IX	512K	1- 10 MB	8088	6	~$ 5,200
	PDP11/23 Vnx	Venix	256K	2- 5 MB	F-11	?	~$ 32,000
	IBM XT Venix	Venix 86	512K	1- 40 MB	8088	6	~$ 5,200
	SCI 1000	UNIX Sys III+	640K	1- 10 MB	80186	8	~$ 6,500
	Omnibyte	Idris	384K	1- 20 MB	68000	6	~$ 8,500
	m16b 384 15M	Xenix 1.3.0	384K	1- 15 MBi	68000	6	~$ 7,500
	PDP11/23 V7	UNIX V7	256K	2- 10 MB	F-11	?	~$ 32,000
	DEC Pro/350	Venix	256K	1- 5 MB	F-11	?	~$ 8,500
	Apple Lisa	UNIX Sys III+	1024K	1- 5 MB	68000	6	~$ 9,190
#2	CT MinF 1.5M	UNIX Sys V	1536K	1- 36 MB	68010	8	~$ 12,000
	CT MinF 512K	UNIX Sys V	512K	1- 20 MB	68010	8	~$ 7,000
	m16b 512 15M	Xenix 1.3.0	512K	1- 15 MBi	68000	6	~$ 8,000
#3	AT&T 3B2/300	UNIX SYS 5	2048K	1- 40 MBi	32000	12	~$ 13,000
	m6000 1M 15M	XENIX 3.0	1024K	1- 15 MB	68000	8	~$ 7,343
	m6000 1M 35M	XENIX 3.0	1024K	1- 35 MB	68000	8	~$ 8,343
	IBM AT Xenix	XENIX 3.0	1536K	1- 20 MBi	80286	6	~$ 13,000
	TANDY 2000	XENIX 3.0	512K	1- 10 MBi	80186	8	~$ 5,500
	m16b 768 12M	Xenix 1.3.2	768K	1- 12 MB	68000	6	~$ 8,700
	m16a 512 12M	Xenix 1.3.2	512K	1- 12 MB	68000	6	~$ 9,191

The Final Outcome First

Detailed documentation for the eight different benchmark tests follows this summary page. After looking at these combined scores for all machines tested, some of the detailed explanations of the individual benchmarks might be of further interest. For example, to relate the "smaller" computers with the classical MIPS (Million Instructions Per Second) ratings of the larger machines, the graph for the CPU Loop Test allows a direct reading in MIPS, as compared to a known value.

When combining the percentile scores of all benchmark tests, the Multi-Tasking and Disk Read benchmarks were given double weight in the final result. This is mainly because performance in these two specific areas is more critical to good performance in a high load multi-user environment. In the following graph, a perfect score of "100%" would result if a system ranked number one (had the lowest elapsed time) in every test. The lowest possible combined performance score approaches "0.00%" as a limit. For more detail about scoring, see the explanation later in this paper. It is important to note that this combined score is for one to six users only.

Machine Tested	Weighted Score	Bar Graph of Combined Overall System Performance
VAX 11/780	76.22%	***
AT&T 3B2/300	74.17%	**
Masscomp	62.14%	**
IBM AT Xenix	61.02%	**
CT MinF 1.5M	57.64%	***
Sun 2/120	54.57%	**
VAX 11/750	53.95%	**
CT MinF 512K	50.74%	***************************************
PDP 11/70	47.78%	*************************************
Altos 986	45.65%	***********************************
m6000 1M 35M	43.14%	*********************************
m6000 1M 15M	40.72%	********************************
Tandy 2000	27.89%	*********************
SCI 1000	25.16%	*******************
m16b 768 12M	24.37%	*******************
m16b 384 15M	21.34%	****************
m16b 512 15M	20.86%	****************
Apple Lisa	19.15%	***************
m16a 512 12M	17.86%	**************
IBM XT Venix	13.01%	**********
IBM XT PC/IX	12.67%	*********
PDP11/23 Vnx	12.16%	*********
PDP11/23 V7	11.83%	*********
Omnibyte	11.74%	********* (data was incomplete)
DEC Pro/350	11.13%	********

In summary, the VAX 780 is rated lower than would be expected, mainly because of the lower performance disk drives attached to it. With these slower drives attached to the VAX, machines such as the AT&T 3B2/300 computer (with a high performance disk) can come close to matching its performance scores in the 1-6 user range. Most all of the higher performance machines have high performance drives attached. This leads directly to the conclusion that the disk drive attached to a computer is the single most important factor contributing to system performance. Given that very fast processors can be attached to almost any disk drive, it makes no sense to attach low speed disk drives to high performance processors.

Performance Benchmark for Multiple Concurrent Tasks

This is a graph of the percentile scores of the concurrent processes benchmark test. It shows the weighted multi-user performance of each system. Since the scores are more heavily weighted toward running a larger number of concurrent background tasks (see Scoring Explanation), the score shows the relative performance of each machine with any combination of **up to six** users active. The graph is sorted by the percentile score. A perfect score of 100% would have resulted from having the smallest elapsed time for all 6 benchmark runs. Since some systems did not (could not) successfully complete all six tests, they were "awarded" a very large real time (99,999 seconds) for any failures. The number is to make the combining arithmetic work (works better than alpha "crashed").

The percentile scores are based on the real execution time, as provided by a UNIX system program. Using the real time for this test most closely approximates what the user sees at his terminal. The background process includes sorting, piping, and listing operations on a small data file provided in the test itself. These functions were deemed representative of the tasks that an "average" user attempts for the majority of his session at a terminal. This benchmark is a good indicator of how a system responds with one to six users active. The normal customer buying Radio Shack computers, is probably only interested in this size range, so it was not judged important to try to get complete data for more than six concurrent users.

Machine	Score	Bar Graph of Weighted Multi-User Performance (1-6 Users) 100%
VAX 11/780	99.22%	***
VAX 11/750	88.79%	***
AT&T 3B2/300	83.09%	***
IBM AT Xenix	82.22%	***
CT MinF 1.5M	81.18%	**
Sun 2/120	80.64%	**
Masscomp	80.59%	**
PDP 11/70	78.58%	***
m6000 1M 35M	68.18%	**
m6000 1M 15M	56.07%	*********************************
Altos 986	51.34%	******************************
m16b 768 12M	42.33%	*************************
CT MinF 512K	38.80%	***********************
Tandy 2000	27.53%	****************
m16b 512 15M	18.63%	***********
m16b 384 15M	17.09%	**********
m16a 512 12M	16.50%	*********
IBM XT PC/IX	14.90%	********
SCI 1000	13.46%	********
PDP11/23 V7	12.63%	*******
Apple Lisa	6.31%	***
IBM XT Venix	6.24%	***
DEC Pro/350	3.30%	*
PDP11/23 Vnx	2.83%	*
Omnibyte	#####	(data not available)

In general, the systems with the most memory performed best in this benchmark. This is not unexpected. The amount of free memory for users to share is known to be the most important factor in multi-user performance, along with the speed of the disk drives (average access time). The speed of the CPU is the third most important factor in multi-user performance, along with the general efficiency of the operating system.

A Graph of Single Background Task Performance

This graph of the single tasking performance of the tested systems is included for the interest of those who are looking for a smaller system performance guide. The benchmark times for running a single background process (taken from the first step in the multi-user test above) were used to compute final percentile scores. The actual elapsed times are shown in another column on the graph. This "Time" column is the real time provided by a UNIX system program. The percentile scoring was done in the same manner as the rest of the benchmarks (minimum elapsed time is 100% with all other times being scored relative to this minimum).

The time column (in seconds) also shows that the background process used for timing the multi-user benchmarks was not a trivial task for any of the machines tested.

Machine	Time	Score	Bar Graph of Single Background Task Performance 100%
Sun 2/120	3.60	100.00%	**
AT&T 3B2/300	4.00	90.00%	**
Masscomp	4.20	85.71%	***
VAX 11/780	4.30	83.72%	**
VAX 11/750	4.30	83.72%	**
IBM AT Xenix	4.46	80.72%	**
PDP 11/70	5.00	72.00%	***
CT MinF 1.5M	6.26	57.51%	**********************************
Altos 986	6.30	57.14%	**********************************
m6000 1M 35M	6.44	55.90%	*********************************
CT MinF 512K	7.80	46.15%	***************************
Tandy 2000	8.20	43.90%	**************************
m6000 1M 15M	9.20	39.13%	***********************
IBM XT PC/IX	10.60	33.96%	********************
m16b 768 12M	13.70	26.28%	*************** (Xenix reconfigured for size)
PDP11/23 Vnx	14.00	25.71%	***************
IBM XT Venix	15.00	24.00%	**************
SCI 1000	15.10	23.84%	**************
m16b 512 15M	16.50	21.82%	*************
m16b 384 15M	20.00	18.00%	**********
PDP11/23 V7	22.30	16.14%	*********
m16a 512 12M	22.50	16.00%	*********
DEC Pro/350	26.00	13.85%	********
Apple Lisa	38.10	9.45%	*****
Omnibyte	#####	#####	(data not available)

The ordering in this graph is very different from the weighted graph for multi-user benchmarks running one to six background processes. This graph mainly shows the folly of attempting to predict multi-user performance based on the timings of single-tasking. The graph on the next page is of some interest in looking at the overall performance ratings of single user systems, as compared to the weighted multi-user data used on the first graph above. The total available memory is less of a factor in this benchmark than in the multi-user tests, giving the disk speed even more relative importance.

Computed Overall SINGLE-USER Performance

This ordering of this graph is different from the multi-user overall performance graph (Page 2) only in the fact that the times for running one background process were used. The timings for 2-6 background proccesses were omitted totally. The final score here gives double importance to the Disk Read benchmark times, but NOT to the single background process time. Otherwise, the scoring is the same as it was for the multi-user overall performance graph.

Machine	Score	Bar Graph of Overall Single User Performance
AT&T 3B2/300	73.95%	***
VAX 11/780	71.94%	**
Masscomp	60.65%	***
IBM AT Xenix	58.50%	***
Sun 2/120	53.82%	***
CT MinF 512K	52.89%	**
CT MinF 1.5M	52.40%	**
VAX 11/750	49.52%	**
Altos 986	45.66%	*************************************
PDP 11/70	43.63%	***********************************
m6000 1M 35M	38.99%	*******************************
m6000 1M 15M	37.13%	******************************
Tandy 2000	29.75%	************************
SCI 1000	27.61%	**********************
ml6b 384 15M	21.92%	*****************
ml6b 512 15M	21.46%	*****************
Apple Lisa	20.93%	****************
ml6b 768 12M	20.59%	****************
ml6a 512 12M	17.96%	**************
PDP11/23 Vnx	15.74%	************
IBM XT Venix	15.74%	************
IBM XT PC/IX	14.54%	***********
DEC Pro/350	13.17%	**********
Omnibyte	13.04%	********** (data was incomplete)
PDP11/23 V7	12.13%	*********

The ordering of the systems as SINGLE USER machines is different from the multi-user ordering. In this graph, one can detect that the speed of the disk drive and the speed of the processor are the two most important rating factors. This is opposed to the size of real memory being more important than processor speed in multi-user system performance.

UNIX Kernel Efficiency Benchmark

This graph compares the performance of UNIX "pipes" in a direct manner. The results can be very misleading if the real time to test completion is used, since a very slow disk drive can skew the real time. For this reason, this benchmark was computed using only the "system time" value as supplied by UNIX. The system time is a direct measure of the kernel efficiency. Pipe efficiency can be a decisive factor in overall UNIX system performance, since pipes are very commonly used in UNIX programs and scripts. In the original BYTE article, this test had to be modified slightly to run on Omnibyte's Idris operating system (One wonders if it helped?).

Machine	Time	Score	Bar Graph of Kernel Efficiency (Pipes Test) 100%
VAX 11/780	1.20	100.00%	**
CT MinF 512K	1.40	85.71%	***
CT MinF 1.5M	1.50	80.00%	**
VAX 11/750	2.10	57.14%	**********************************
IBM AT Xenix	2.26	53.10%	*******************************
m6000 1M 15M	2.50	48.00%	****************************
m6000 1M 35M	2.60	46.15%	***************************
AT&T 3B2/300	2.70	44.44%	**************************
Tandy 2000	2.70	44.44%	**************************
Masscomp	2.80	42.86%	*************************
Altos 986	2.80	42.86%	*************************
Apple Lisa	3.00	40.00%	************************
SCI 1000	3.10	38.71%	***********************
PDP 11/70	3.40	35.29%	*********************
m16b 384 15M	3.40	35.29%	*********************
Sun 2/120	3.70	32.43%	*******************
m16a 512 12M	3.79	31.66%	******************
m16b 512 15M	3.80	31.58%	******************
m16b 768 12M	4.51	26.61%	***************
IBM XT Venix	7.30	16.44%	*********
IBM XT PC/IX	7.60	15.79%	*********
PDP11/23 Vnx	9.50	12.63%	*******
PDP11/23 V7	10.70	11.21%	******
DEC Pro/350	13.80	8.70%	*****
Omnibyte	30.40	3.95%	**

In studying this graph, it seems that special attention has been paid to pipe efficiency in the newer versions of Xenix, especially those on the Tandy and IBM machines. If the DEC Pro/350 has not already been pulled from the market, we can expect it soon, the machine looks to be in deep trouble. The Convergent Technology Mini Frames appear to have had special attention paid to making an efficient UNIX port, the version of UNIX System 5 runs on the Motorola 68010 processor.

Benchmark Efficiency in Making System Calls

This graph shows a direct performance measurement of calling the operating system to perform a service. The elapsed real time to make multiple calls is used as the basis for computing the percentile scores. This test gives a direct measure of the elapsed time required to make a system call and then return to the caller. The test program compounds the kernel overhead involved in executing a system call. Making a system call involves an interrupt trap to the kernel or into supervisor mode, performing the desired function and returning. There are almost no programs written which do not make system calls. The most common system calls are for input/output services, such as open, close, read, and write.

Machine	Time	Score	Bar Graph of System Call Efficiency 100%
VAX 11/780	4.80	100.00%	**
Masscomp	6.30	76.19%	***
Sun 2/120	6.80	70.59%	**
VAX 11/750	7.00	68.57%	***
IBM AT Xenix	7.11	67.51%	**
CT MinF 1.5M	7.60	63.16%	*************************************
CT MinF 512K	7.60	63.16%	*************************************
PDP 11/70	8.00	60.00%	************************************
AT&T 3B2/300	9.30	51.61%	******************************
m6000 1M 15M	9.80	48.98%	*****************************
m6000 1M 35M	9.80	48.98%	*****************************
Apple Lisa	10.50	45.71%	***************************
Altos 986	11.00	43.64%	**************************
Tandy 2000	14.20	33.80%	********************
m16b 768 12M	14.70	32.65%	*******************
m16b 384 15M	15.00	32.00%	*******************
m16b 512 15M	15.10	31.79%	*******************
m16a 512 12M	16.40	29.27%	*****************
IBM XT Venix	20.50	23.41%	**************
Omnibyte	21.30	22.54%	*************
PDP11/23 Vnx	24.00	20.00%	************
SCI 1000	26.20	18.32%	**********
DEC Pro/350	33.30	14.41%	********
PDP11/23 V7	36.50	13.15%	*******
IBM XT PC/IX	39.80	12.06%	*******

The large jump from the VAX to the next system in line is caused by there being special context switching instructions available to the VAX systems programmer (in addition to the raw CPU speed available). This fact lets the efficiency of system calls be maximized (minimizing the time required). Otherwise, there are several systems which do a poorer job of this system call context switching. One interesting note here is that the Intel 80286 in the IBM AT seems to be more efficient at interrupt handling than the AT&T WE32000 chip. This is not the only tradeoff involved in designing a system, so too much emphasis should not be placed on this one datum.

Efficiency in Making Function Calls from "C" programs

The graph is sorted by the difference in the user time field of the function call results. The difference is the measured time between the elapsed time required to run a program with a function call and then the elapsed time to run the same program without the function call.

Function calls are pervasive in any good programming system. They are commonly used routines which need not and should not be rewritten in every module of a system. The overhead in using them is the time to "stack" the current status of the calling program, set up and execute the function, and then return to the caller with the results of the function. This test uses an "empty" function for timing purposes, one that immediately returns to the caller and does not perform any other work. When the function call is not made in the second case, the main program uses the same amount of time in its execution. All of this makes the observed time difference equal to the actual overhead in setting up and calling functions from a program.

Machine	Time	Score	Bar Graph of Function Call Efficiency 100%
Altos 986	0.40	100.00%	**
AT&T 3B2/300	0.42	95.24%	***
CT MinF 512K	0.70	57.14%	**********************************
CT MinF 1.5M	0.80	50.00%	******************************
Sun 2/120	0.80	50.00%	******************************
Masscomp	0.90	44.44%	**************************
m6000 1M 15M	0.93	43.01%	*************************
m6000 1M 35M	0.93	43.01%	*************************
VAX 11/780	1.00	40.00%	************************
PDP 11/70	1.00	40.00%	************************
SCI 1000	1.20	33.33%	*******************
Apple Lisa	1.30	30.77%	******************
IBM AT Xenix	1.30	30.77%	******************
m16b 384 15M	1.40	28.57%	*****************
m16b 512 15M	1.40	28.57%	*****************
m16a 512 12M	1.47	27.21%	****************
m16b 768 12M	1.48	27.03%	****************
Tandy 2000	1.50	26.67%	***************
VAX 11/750	1.70	23.53%	**************
Omnibyte	1.70	23.53%	**************
IBM XT Venix	2.80	14.29%	********
PDP11/23 Vnx	3.30	12.12%	*******
DEC Pro/350	3.50	11.43%	******
PDP11/23 V7	3.60	11.11%	******
IBM XT PC/IX	4.70	8.51%	*****

The range of time differences here is much smaller than in some other sections of the benchmark tests. Since function call efficiency is being measured, notice that the IBM PC AT and the AT&T 3B2/300 have reversed their positions from the system call benchmark. The MC68000 machines do very well in this test, but some credit must go to the C compiler on some machines for generating good optimized object code.

Computationally Generating Prime Numbers

No benchmark series would be complete without some form of a Sieve of Eratosthenes to compute some quantity of prime numbers. This graph of a sieve program is sorted by the percentile scores derived from the elapsed real time field of the sieve program results. This is mostly a compute bound job measuring both processor speed and processor to memory speed, but which also measures the C compiler's ability to generate good efficient object code.

Machine	Time	Score	Bar Graph of Prime Number Sieve Program 100%
VAX 11/780	1.70	100.00%	**
PDP 11/70	2.30	73.91%	**
CT MinF 1.5M	2.60	65.38%	***************************************
CT MinF 512K	2.60	65.38%	***************************************
VAX 11/750	2.70	62.96%	*************************************
Masscomp	2.80	60.71%	************************************
IBM AT Xenix	3.04	55.92%	*********************************
m6000 1M 15M	3.20	53.13%	*******************************
m6000 1M 35M	3.20	53.13%	*******************************
Altos 986	3.30	51.52%	******************************
AT&T 3B2/300	3.90	43.59%	**************************
Tandy 2000	4.10	41.46%	************************
SCI 1000	4.40	38.64%	***********************
Sun 2/120	5.10	33.33%	*******************
m16b 768 12M	5.30	32.08%	*******************
PDP11/23 Vnx	5.50	30.91%	******************
PDP11/23 V7	5.80	29.31%	*****************
m16b 384 15M	6.00	28.33%	****************
Apple Lisa	6.10	27.87%	****************
DEC Pro/350	6.30	26.98%	****************
m16b 512 15M	6.60	25.76%	***************
Omnibyte	7.00	24.29%	**************
m16a 512 12M	7.30	23.29%	*************
IBM XT PC/IX	8.20	20.73%	************
IBM XT Venix	9.00	18.89%	***********

The fact that the Tandy Model 6000 has exactly the same score for machines which differ only in the disk drive attached, means that this benchmark is a true measure of internal system performance of the processor, bus, and memory. The graph would probably be valid for all types of compute bound jobs. Since this test is compute bound, the processor speed should be of primary importance.

Using this graph in conjunction with the CPU Loop test (Page 12) shows the relative memory speed and bus efficiency of the systems measured. The Loop test shows the outright speed of the CPU and this test has memory speed factored into it since the prime numbers are stored into memory as well as all of working loop counters, etc.).

Benchmark Test of Disk Write Performance

The graph is sorted by the percentile score computed from the elapsed real time, also shown, of the Disk Write test program. The graphed time is the time required to write 512 records of 256 bytes each onto a disk file. The UNIX operating system is known to be a very disk intensive system and this test gives a general feel for the overall performance of the system. The overall performance is affected by the rotational delay of the disk drive, the average access time, the disk sector interleave factor, and the disk cache buffering algorithms.

One would do well to remember that each of the several manufacturers represented here has his own definition of write "verify". The variations are numerous but include: no read after write at all; read after write on directory records only; and read after write on all disk sectors. Because there are so many possible variations of implementation of "write verify" which can affect the benchmark performance, the reader should be aware that the disk write verify condition was not known or controlled for all machines tested. All of the model 16's had write verify turned off for these benchmarks.

Machine	Time	Score	Bar Graph of Disk File Write Performance 100%
AT&T 3B2/300	1.50	100.00%	**
Masscomp	1.70	88.24%	**
Sun 2/120	1.80	83.33%	***
VAX 11/780	2.00	75.00%	***
CT MinF 512K	2.10	71.43%	**
CT MinF 1.5M	2.30	65.22%	***************************************
IBM AT Xenix	2.31	64.94%	**************************************
VAX 11/750	3.00	50.00%	******************************
Altos 986	3.50	42.86%	*************************
PDP 11/70	4.00	37.50%	**********************
SCI 1000	4.30	34.88%	********************
m6000 1M 35M	4.60	32.61%	*******************
m6000 1M 15M	4.70	31.91%	*******************
Tandy 2000	5.10	29.41%	*****************
IBM XT Venix	7.00	21.43%	************
DEC Pro/350	7.70	19.48%	***********
PDP11/23 Vnx	8.00	18.75%	***********
m16b 384 15M	8.00	18.75%	***********
m16b 512 15M	9.10	16.48%	*********
m16b 768 12M	10.90	13.76%	********
IBM XT PC/IX	11.60	12.93%	*******
Omnibyte	12.30	12.20%	*******
m16a 512 12M	14.40	10.42%	******
Apple Lisa	20.80	7.21%	****
PDP11/23 V7	22.00	6.82%	****

This graph is mostly a chart of disk drive write speed, factored by the disk caching methodology. One thing to pay particular attention to is the differences in real times on the "m6000 1M 15M" and the "m16b 512 15M". There are several improvements in the Tandy model 6000 for disk access, these can be seen directly since the two disk drives are identical. This graph could well be a graph of all of the different disk drives attached to a single system since the scores are relative. AT&T has claimed a super efficient disk caching scheme for UNIX System 5.

Benchmark of Random Disk Record Read Performance

The graph is sorted by the percentile score computed from the elapsed real time of the Disk Read benchmark program. This benchmark test program reads "random" disk file records from the file created by the disk write test. The graph is a direct measure of the random disk read speed of each system. THE TIMES FOR the Masscomp and the Omnibyte machines were not available as true benchmark timings, so they were ESTIMATED based on timings from the disk write benchmark. ONLY these two numbers should be considered as best guesses.

Machine	Time	Score	Bar Graph of Disk Read Performance (Random records) 100%
AT&T 3B2/300	1.95	100.00%	**
IBM AT Xenix	2.55	76.47%	***
Masscomp	3.60	54.17%	********************************
Sun 2/120	4.90	39.80%	***********************
CT MinF 1.5M	7.13	27.35%	****************
Altos 986	7.30	26.71%	****************
CT MinF 512K	7.68	25.39%	***************
VAX 11/780	8.00	24.38%	**************
VAX 11/750	8.00	24.38%	**************
SCI 1000	9.10	21.43%	************
PDP 11/70	9.50	20.53%	************
m6000 1M 35M	10.20	19.12%	***********
m6000 1M 15M	10.50	18.57%	***********
Tandy 2000	12.70	15.35%	*********
IBM XT PC/IX	20.70	9.42%	*****
m16b 512 15M	20.70	9.42%	*****
m16b 384 15M	22.00	8.86%	*****
Omnibyte	25.00	7.80%	****
IBM XT Venix	25.60	7.62%	****
DEC Pro/350	28.00	6.96%	****
PDP11/23 V7	32.70	5.96%	***
PDP11/23 Vnx	33.70	5.79%	***
Apple Lisa	44.50	4.38%	**
m16b 768 12M	54.30	3.59%	**
m16a 512 12M	55.40	3.52%	**

The main performance factor for disk drives in this test is the "Average Access Time". The average access time number for each of the disk drives could almost be used as the direct scale on the above graph without changing the relative ratings to any marked degree. Notice that the VAX 780, a "1 MIPS" processor, has placed very low in the graph because the average access time of the disks attached to it does not compare favorably with some of the disk drives used with the "smaller" machines. There are some smaller variations in this graph that are attributable to the size of the disk "cache" in number of buffers available for disk reads. The "average access time" is so large in comparison to even the slowest CPU that the system overhead has almost no effect on the final outcome.

Anyone who is trying to build a high performance UNIX system should attach disk drives that have the lowest average access time available in the market. All other money spent pales in comparison to this often overlooked major factor. Notice in the graph that AT&T and IBM already know and use this fact in their small micros.

The CPU "Power" Ratings

The graph is sorted by the percentile scores derived from the elapsed real time field of the Loop test results. The program tests the incrementing of long integers in a CPU register. It is totally compute bound and directly compares the processor speeds of the tested systems. Since the VAX 11/780 is rated as a "1 MIPS" processor (according to DEC's local sales office), the percentage scale turns out to be directly readable as the "MIPS" rating of each CPU tested. As a point of information, IBM's latest mainframe announcement, the "Sierra" (dyadic) machines, are reportedly designated as 29 MIPS. IBM has also promised a 50-80 MIPS machine in the near time frame.

Machine	Time	Score	Bar Graph of CPU Loop Test (0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0) 100%
VAX 11/780	2.60	100.00%	**
VAX 11/750	5.10	50.98%	******************************
AT&T 3B2/300	6.40	40.63%	************************
Masscomp	6.60	39.39%	***********************
CT MinF 512K	7.18	36.21%	*********************
CT MinF 1.5M	7.30	35.62%	*********************
Sun 2/120	7.40	35.14%	*********************
PDP 11/70	7.90	32.91%	*******************
m6000 1M 15M	7.90	32.91%	*******************
m6000 1M 35M	7.90	32.91%	*******************
IBM AT Xenix	12.61	20.62%	************
m16b 768 12M	13.20	19.70%	***********
Altos 986	13.30	19.55%	***********
m16b 384 15M	14.00	18.57%	***********
Apple Lisa	14.00	18.57%	***********
m16b 512 15M	14.20	18.31%	**********
SCI 1000	14.50	17.93%	**********
Tandy 2000	15.00	17.33%	**********
m16a 512 12M	15.50	16.77%	**********
Omnibyte	17.00	15.29%	*********
PDP11/23 Vnx	26.00	10.00%	******
DEC Pro/350	26.70	9.74%	*****
PDP11/23 V7	27.40	9.49%	*****
IBM XT PC/IX	32.20	8.07%	****
IBM XT Venix	32.70	7.95%	****
Direct Read MIPS scale			0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0

It is possible to see clearly the different processors (and clock speeds) used in the tested machines. From the slowest Intel 8088 in the IBM XT's to the faster 68000's and up to the 68010's and WE32000 (Used in the AT&T 3B2/300). Note that the 6Mhz 68000 is used in the Model 16b's, the Apple Lisa, and Altos units. The 68000 6MHz processor is equivalent to a rating of about .15 MIPS (a bit more than the 8MHz 80186).

If one REALLY studies this graph, he will discover that the CPU architecture is much less important than the clock speed as a performance factor. One should always choose a processor that has the faster clock speed, followed by an instruction set that makes for the least amount of housekeeping overhead, especially in address spaces. The appreciable thing about this graph is that there are no real surprises. This lends credibility to the other benchmark tests for the same systems.

The "Slope" of Multi-User Performance

In order to try to show the RELATIVE user perception of the systems in this test as multi-user systems, the following graph was developed. To begin with, the "slope" time was computed from each step of 1-2, 2-3, 3-4, 4-5, & 5-6 concurrent tasks. The slope time is the difference in elapsed real time from "n" processes to "n+1" processes running at the same time. After the slope time was computed for each machine, it was averaged. This gave a number representing the "worsening" effect of each multi-user step as it is added. The second number is an average elapsed real time for all six of the concurrent background shell tests. The two numbers were next multiplied to compound the "worsening" effect of adding users. After this multiplication, the meaningless numbers were converted to percentile scores (the lowest being 100%) for graphical purposes.

Looking at the graph should reveal a truer picture of any of the machines as a user perceived multi-user system. The obvious dichotomy in the graph should indicate which machines are truly multi-user machines and which are truly single-user machines. This graph tells nothing about performance in either environment, it only indicates those machines which can be considered as "multi-user" systems and which cannot.

Machine	Score	Bar Graph of Six User "Slope" Performance	100%
VAX 11/780	100.00%	**	
VAX 11/750	79.78%	***	
CT MinF 1.5M	79.56%	***	
AT&T 3B2/300	65.19%	***************************************	
PDP 11/70	64.31%	**************************************	
IBM AT Xenix	64.30%	**************************************	
Masscomp	57.08%	**********************************	
Sun 2/120	54.08%	********************************	
m6000 1M 35M	51.80%	*******************************	
m6000 1M 15M	39.15%	***********************	
m16b 768 12M	23.26%	************* (questionable)	
Altos 986	15.60%	*********	
CT MinF 512K	11.11%	******	
Tandy 2000	4.20%	** (single-user)	
m16a 512 12M	2.19%	*	
m16b 512 15M	2.13%	*	
m16b 384 15M	2.03%	*	
SCI 1000	1.24%		
PDP11/23 V7	1.18%		
IBM XT PC/IX	1.11%		
IBM XT Venix	1.05%		
PDP11/23 Vnx	0.74%	(less than one user)	
Apple Lisa	0.32%		
DEC Pro/350	0.23%		
Omnibyte	#####	(data not available)	

Notice that the systems are almost ordered by real memory size within average access time of the disk drive. This aids the believability of this graph as a good indicator of multi-user performance. One does find it hard to understand how the 1024K memory Apple Lisa with a 6MHz 68000 CPU can be made to perform so poorly.

All of the graphs that have a "Time" column in this report were computed in the same manner. First a choice was made of the lowest time for all systems in a particular run. Following this choice of the least time, each of the numbers for each of the other systems was divided into it. This has the effect of making the system with the lowest time a 100 percentile score and all of the other systems are percentile scores of lesser magnitude. This method also has the effect of scaling the graphs to the size of the paper.

There is a graph on Page 3 of "weighted multi-user performance." The scores on this page were computed based on the idea that the more shells that can be run more quickly, the better is the machine's performance. The score was computed by taking each of the runs for a different number of background processes and treating it as a separate run. These separate runs were scored in the same way as the other runs were scored, actual time divided into least time of all systems for that run. After each of the runs (1-2-3-4-5-6) had percentile scores (pct#1, pct#2, pct#3, pct#4, pct#5, pct#6) for each system, it was relatively easy to then combine them. The final scores for each system are computed with the formula:

(1*pct#1 + 2*pct#2 + 3*pct#3 + 4*pct#4 + 5*pct#5 + 6*pct#6)/21

These computed scores for each system were used as the combined "weighted percentile scores", shown in the "Score" column on Page 3. It was chosen to weight them in this manner to give more importance to running more concurrent background tasks. The graph on Page 4 shows only the time and the percentile score for running one background task. This graph was drawn to show the relative initial capability of the systems before several concurrent tasks are started.

The combined overall score on Page 2 (Final Outcome) is very simply the average of all of the percentile scores from all of the different runs (treating the combined multi-user score as ONE score, not six). Because the performance of the disk read (UNIX is very disk intensive) and the combined multi-user scores (concurrent tasks ability) were thought to be more important to the final outcome, each of these two (combined multi-user & disk read) percentile scores were added twice into the average. This average overall score is the one graphed on Page 2.

There is a graph on Page 5 of the Combined Single-user Performances of the systems. This final score was computed by replacing the combined multi-user score with the score for running one background task. All runs were then averaged with the disk read score being added twice into the average.

The multi-user "slope" chart on Page 13 was scored as explained on that page.

In Conclusion

It would be unjust to look at the Final Outcome chart on Page 2 and say that the AT&T 3B2/300 is "close to the VAX 780" in performance. The "real world" factor that must be considered is the TOTAL AMOUNT OF DISK SPACE AVAILABLE. Assuming that the AT&T system has an equivalent 512 MB of disk storage, then it would be fair to say that FOR UP TO SIX USERS there is little difference between the two. Many similar comparisons can be drawn from the final outcome, but they are all moot points if the total amount of disk space available is not adequate for the application. What, then, if one needs support 10 users instead of six?

It is fair to say that, as a rule of thumb, a good multi-tasking operating system will use 10 MB of storage on disk. Using a similar rule, each user who "logs in" will also need 10 MB of disk space. A single user system (one user active at a time) that is used by six people (who log in at different times) or a multi-user system having a total of six users who log in and can be active at one time are both 70 MB sized systems so far as disk space is concerned. Other space requirements for large data files should be added but it is fair to say that a six user system should have roughly 70 MB of disk space available to be of useful size. Nowhere in any benchmark can it be observed that this is true. One should be sure to consider this fact when selling or buying systems.

For those who wish to "emotionally" continue the old Motorola vs. Intel processor wars, this paper is a fertile ground for finding substantiating "facts", without regard for context or overview. As a matter of fact, the only difference that has any real meaning is "How easy is it to program?". The simpler, the better. Benchmarkwise, the only factor which can be considered as conclusive is the clock rate of the CPU. One area which does seem to support the arguments is the CPU loop test which shows that the 6MHz 68000 is roughly the same speed as the 8MHz 80186. I surmise that there are probably about 2MHz worth of wait states in the 80186 architectures which cause this to be observable. Otherwise the "bank switchers" (Intel) are really more inferior than one could reasonably predict.

As far as Tandy's offerings go, the Model 6000 is a significant improvement over the Model 16b. Recommendations for multi-user systems should be limited to the 35 MB disk drives. On the average, there is enough disk space for TWO users plus the operating system on a single 35 MB disk and 512K of RAM. A 2-4 disk (35 MB) system, with 1 MB RAM, should be well sized for 4-8 users. Possibly 2MB of RAM might allow 10 users to be active on four disks, so long as all are not compute bound. These statements are not really true for the model 16b. The Tandy 2000 should be sold as a single-user system only, however, there needs to be some attention paid to being able to add at least 10 MB worth of disk space for an average system user. The internal 10 MB is about right for the operating system plus utility, log, and work files.

Either the AT&T 3B2/300 or the Convergent Technologies systems could be used as "upgrades" from the Model 6000. Before this can happen, each system needs to find enough disk and memory space add-ons to be able to handle the extra user load. An 8-10 user system should have about 2-3 MB of RAM and about 100-256 MB of disk to really be useful commercially. Anything less than this will not be easy to use or to sell. The average access time of any disks added should be MINIMIZED for the best performance as a multi-user system.

I believe that the "bang for the buck" (price/performance) over the last few years has demonstrated considerable improvement. The old DEC PDP 11/23 that has been in the field (to the tune of ~100,000 systems installed [Datapro]) for years shows the poorest performance for the money, while the newest machines from Tandy, IBM, and AT&T all show excellent performance for the price (see next page).

A Graph of Price/Performance

This graph was derived for looking at the price versus the performance of the tested systems. Remembering that this chart is for ONE TO SIX USER systems only, it is interesting to note that the smaller (slower) machines seem to do quite well in the analysis. The graph was computed by taking the approximate retail price from Page 1 and multiplying it by the quantity (1-final score). The final score is the one appearing on Page 2. This derived number is labeled "Cost" in the graph. The percentiles were assigned based on the minimum "cost" of all systems.

All of the retail prices are not known for absolute certain and everyone knows that prices are discounted by different amounts from different sources. Taking all of this into account, the "ballpark" price is not too far off of the manufacturer's "List Price".

Machine	Cost	Score	Bar Graph of Price/Performance	100%
AT&T 3B2/300	3358	100.00%	**	
CT MinF 512K	3448	97.39%	**	
Tandy 2000	3966	84.67%	**	
m6000 1M 15M	4353	77.15%	**	
IBM XT Venix	4523	74.24%	**	
IBM XT PC/IX	4541	73.95%	**	
m6000 1M 35M	4744	70.79%	**	
SCI 1000	4865	69.03%	***	
IBM AT Xenix	5067	66.27%	***************************************	
CT MinF 1.5M	5083	66.07%	***************************************	
ml6b 384 15M	5899	56.92%	**********************************	
ml6b 512 15M	6331	53.04%	*******************************	
Altos 986	6522	51.49%	******************************	
ml6b 768 12M	6580	51.03%	******************************	
Apple Lisa	7430	45.20%	***************************	
Omnibyte	7502	44.76%	**************************	
ml6a 512 12M	7549	44.48%	**************************	
DEC Pro/350	7554	44.45%	**************************	
Sun 2/120	11358	29.57%	*****************	
Masscomp	11359	29.56%	*****************	
PDP11/23 Vnx	28107	11.95%	*******	
PDP11/23 V7	28215	11.90%	*******	
VAX 11/780	83230	4.03%	**	
VAX 11/750	85188	3.94%	**	
PDP 11/70	93991	3.57%	**	

It is not unreasonable to preclude the use of the "large" DEC machines when talking about 1-6 user systems. It is interesting that the price/performance of the slower machines is so good. Notice that this graph measures price/performance and not the absolute speed of the system.