Characterization of Resource Availability on PlanetLab Nodes for Executing Migratory Tasks

Vinit Padhye and Anand Tripathi *
University of Minnesota, Minneapolis MN 55455

Abstract

We present here the results of our study to monitor and characterize the available resource capacities on PlanetLab nodes for scheduling and dynamic relocation of migratory tasks. The available resource capacities on PlanetLab nodes can fluctuate significantly in a short time, and a node selected to execute a task may become unsuitable for it within a short time ranging from a few minutes to an hour. One way to overcome this problem is relocate the task to another node. This requires identifying the set of nodes that are mostly likely to meet the requirements of the task for a long duration. For this purpose we have developed a system for monitoring a collection of nodes for their available resource capacities. Using this system we have conducted a study to characterize the distribution of the number of nodes that meet a given resource capacity requirement, and the distribution of the duration for which a node satisfies the requirement.

1 Introduction

The available resource capacities on PlanetLab [1] nodes can fluctuate significantly as shown by the study presented in [3]. It is shown there that the available resource capacity at a node may change significantly within an hour. A node selected to execute a task with some given resource requirements can become unsuitable for hosting it in the near future due to the changes in the available resource capacity on that node. In [3] the authors argued for the need of supporting dynamic relocation of tasks based on resource availability. In an earlier work [6] we developed mechanisms for autonomic relocation of tasks in the PlanetLab environment for building resource-aware migratory services using the mobile agent based programming model [5]. We implemented mechanisms for a task, programmed as a mobile agent, to monitor its host execution environment and autonomically migrate to another host if the available resource capacity at the current host falls below some required threshold. The mechanisms for task migration in our environment have been presented in [6], and therefore we do not discuss them here.

For scheduling or relocation of a task, we need to identify the potential target nodes that satisfy the task's resource requirements. We are interested in finding those target nodes that are most likely to satisfy a task's requirement for a long duration, thus minimizing the number of migrations. We refer to the set of potential target nodes satisfying a given resource capacity requirement as its *eligibility set*. The *eligibility period* of a node is defined as the contiguous period for which it remains in the eligibility set. The expected duration for which a node selected randomly from the eligibility set, at an arbitrary point in time, would satisfy a task's requirement is half of the expected value of the eligibility period for that requirement.

The requirements of a task could be stated in terms of CPU capacity, memory, and bandwidth. We need to identify the set of nodes whose currently available resource capacities satisfy the requirements of the task that needs to be scheduled or relocated in the system. To assist in the selection of nodes for the placement and autonomic relocation of tasks in the PlanetLab environment, we have developed a service for periodically monitoring of PlanetLab nodes for their available resource capacities. One can query this service to obtain the set of eligible nodes that satisfy a given resource requirement. This system is used by the migratory tasks to find out the potential target nodes for relocation. We have used this service to study the behavior of PlanetLab nodes in terms of their eligibility periods and eligibility set sizes for a spectrum of resource requirements.

In this paper we present the results of our investigation towards characterization of available resource capacities on the PlanetLab nodes. We were interested in finding the distribution of the eligibility set size for a given resource capacity requirement because it determines the probability of finding a suitable target host when a task is to be scheduled or relocated. The expected value of the eligibility set size is also an indicator of the average number of tasks of a given resource capacity requirement that can be scheduled in the system. The distribution of eligibility periods indicates how long a randomly selected node is likely to meet the given requirement of a task. We also investigated different measures for identifying nodes for inclusion in the eligibility set. Specifically, we examined the impact of taking into account the recent resource availability profile of a node when considering it for inclusion in the eligibility set. We wanted to determine how such profiling affects the expected eligibility period.

In [6] we presented our observations of eligibility periods and eligibility set sizes solely based on CPU capacity requirements. We present here the results of our study of node availability behavior when both CPU and memory requirements are considered together for identifying the eligibility set. Our investigation was driven by several questions. We wanted to understand how the eligibility periods of the nodes and the size

* This work was supported by National Science Foundation grants 0834357 and 0708604

of the eligibility set are affected when both CPU and memory requirements are considered together in comparison to the behavior when only the capacity requirement for CPU or memory is considered in isolation. Second, we wanted to investigate whether any particular dimension of resource requirement, i.e. CPU or memory, dominates. Third, we also wanted to determine how the eligibility set size for a given resource capacity requirement varies over time.

In the next section we present the basic approach and the profiling-based approach for identifying eligible nodes for a given requirement. In Section 3 we present the results of our experiments to evaluate the basic approach using four datasets from different observation periods. Here we compare the eligibility characteristics for CPU, memory, and combined requirements. Section 4 shows how the nodes' availability characteristics are related to their average eligibility periods. The observations from this motivate the need of using a profiling-based approach for selecting the nodes. In Section 5 we show how profiling can significantly increase the expected values of the eligibility periods, but with reduced sizes for the eligibility sets. In the last section we present our conclusions from this study.

2 Monitoring of PlanetLab Nodes for Available Resource Capacities

We present here the approaches that we used and investigated for identifying the set of nodes which satisfy the given resource requirement in terms of available CPU capacity, memory, and bandwidth. We integrated these approaches in a system that we developed for monitoring PlanetLab nodes for their resource utilization. We observed that *CoMon* [4], the node monitoring service on the PlanetLab, could not be used for our purpose directly. This is because of the following reasons. For CPU usage, CoMon provides average values over system-defined monitoring intervals of 1 and 5 minutes. In our experiments, we needed node-level resource utilization data that is collected at a higher frequency (such as every 10-20 seconds) and aggregated to determine statistics over configurable observation intervals. We were also interested in the standard deviation of the measured resource utilization values over the observed periods.

The monitoring service that we developed collects the data about resource consumption of every monitored node by probing its *SliceStat* [4] data every 10 seconds in order to obtain an accurate estimate of the node's resource utilization behavior over time. For selecting the nodes for inclusion in the eligibility set for the given requirement for a resource (such as CPU or memory), we compute the average value of a node's idle capacity for that resource over a window of past 5 minutes. We also take into account the standard deviation of a node's idle capacity over that 5 minutes window to mitigate the effects of fluctuations in the node's idle capacity. For CPU, the idle capacity is expressed in terms of the unused CPU cycles expressed in terms of MHz or GHz.

In [6], we have presented the data in terms of the measures discussed above for CPU requirements only. In this paper, we study the behavior of node availability for memory requirements also as well as CPU and memory conjoined requirements. The first important goal of this study is to compare the distribution of eligibility periods for CPU requirements with that of memory requirements. The key questions in this study are : How the size of eligibility set varies over time? How does the behavior of nodes vary for CPU and memory requirements? Do the nodes show more availability in terms of larger eligibility period and eligibility set size for memory requirements than those for CPU requirements? The second goal of this study is to determine whether the node availability is dominated by either the CPU requirement or the memory requirement, when both CPU and memory requirements are considered together. We also want to see if any relationship between the average eligibility of a node and the fraction of the time it is present in the eligibility set. Finally, another important goal of this work is to investigate how selecting nodes based on their recent profile affects the expected eligibility periods.

We investigated the following two approaches for selecting a node for inclusion in the eligibility set for a given requirement:

Basic Method: If C is the average idle capacity on a node and δ is its standard deviation, then for a given resource requirement R we select the node if it satisfies the following condition:

$$C - 2 * \delta > R \tag{1}$$

A node is dropped from the eligibility set if the idle capacity at that node falls below the resource requirement R. When considering the CPU and memory requirements together for selecting nodes, we select the node only if it satisfies the above condition for both the CPU requirement and the memory requirement. We drop a node from the eligibility set, if either the available CPU capacity or available memory capacity on that node falls below R, the requirement threshold. The eligibility period of a node for a given resource requirement is measured as the time between the node's entry in the eligibility set for that resource requirement and departure from the eligibility set. A node may enter and leave the eligibility set multiple times during the observation period. Thus a node may have multiple eligibility periods. For such nodes, we consider the average value of their individual eligibility periods.

Profiling-based Method: In this approach we wanted to eliminate those node that show highly frequent and significant variation in their available capacity for a given requirement. In this approach we build a *profiled eligibility set* from the basic *eligibility set* constructed using the basic approach presented above. The following rules are used for including a node in the *profiled eligibility set*. The rules use a parameter T, which is a time period value. We set it to 30 minutes in our experiments. A node in the basic *eligibility set* is considered for inclusion in the *profiled eligibility set* if its previously observed eligibility period was greater than 30 minutes, which is the value of the parameter T. If the previously observed eligibility period of the node was less than T, then we include this node in the profiled set only after it has been in the basic eligibility set for the past 30 minutes. When a node is dropped from the basic *eligibility set*, it is also removed from the *profiled eligibility set*. In this approach, the eligibility period of a node is defined as the duration for which it remains in the *profiled eligibility set* for a contiguous interval.

3 Node Selection using the Basic Method

In this section we study the distribution of eligibility periods and eligibility set sizes for a spectrum of resource requirements. In this study we use the basic method for selecting eligible nodes.

In the experiments discussed here we monitored about 200 PlanetLab nodes for their available resource capacities at differ-

CPU	1GHz, 2GHz, 3GHz, 4GHz
Memory	512MB, 1GB, 2GB, 3GB
CPU+Memory	(1GHz + 512MB), (2GHz + 1GB) (3GHz + 2GB), (4GHz + 1GB)

Table 1: Capacity requirements for CPU,Memory and combined (CPU + Memory)

ent periods, ranging from 3 to 7 days at different points during the past six months. Table 1 shows the capacity requirements used in these experiments. The datasets that we collected for our experiments and their observation times are listed in Table 2. During the period for which the Dataset-1, Dataset-3, and Dataset-4 were collected, the monitored PlanetLab nodes were highly loaded while in the case of Dataset-2 they were relatively lightly loaded. All the four datasets correspond to same set of 200 monitored nodes.

Dataset-1	November 18-21, 2009 (75 hours)
Dataset-2	December 1-4, 2009 (97 hours)
Dataset-3	January 22-25, 2010 (96 hours)
Dataset-4	January 26-February 1, 2010 (133 hours)

Table 2: Datasets and their observation times

Figure 1 shows the CDFs of eligibility periods for CPU and memory capacity requirements for the four datasets mentioned above. Table 3 presents statistics such as average, median and standard deviation for eligibility periods and eligibility set sizes for Dataset-1 and Dataset-2.

From Table 3, we observe that typically the median values for the eligibility periods tend to be always less than the average values. The standard deviation also tends to be high, comparable to the average values. This indicates that some nodes tend to exhibit significantly large eligibility periods. This also indicates that the available resource capacities at a node may fluctuate significantly, and there is a large variation of eligibility periods across the nodes. In this table, the *Unique Nodes* column gives the number of nodes that became eligible during the entire duration of the observation. The statistics given in the table for a specific requirement correspond to the unique nodes for that requirement. For example, in case of 2GHz the average eligibility period of 103 minutes is the average of 74 nodes' average eligibility periods. Similarly, in Figure 1(a) the CDF for 2GHZ is the distribution of 74 values for average eligibility periods, whereas for 3GHz the distribution given is for 47 values.

The impact of memory requirements on node eligibility can be understood from the statistics presented in Table 3 and the cumulative distributions given in Figure 1(b,d,f,h). We can observe that nodes show high eligibility periods and eligibility set sizes for memory requirements up to 2GB. However, for Dataset-3 and Dataset-4 we observe that the median values of the eligibility periods are quite small (around 7 to 8 minutes) for 2GB memory requirement. In all the four datasets, we found that very few nodes could satisfy memory requirement of 3GB. This indicates the availability of nodes is high for memory requirements less than 2GB, however, for memory requirements of 2GB or above, the eligibility periods and set sizes can decrease significantly.

The eligibility periods of nodes for combined CPU and memory requirements can be understood by comparing the cumulative distributions of eligibility periods given for the combined requirement in Figure 2 with those for the corresponding CPU and memory requirements shown in Figure 1. We observe that for Dataset-1 and Dataset-2, the distributions of eligibility periods for combined requirements tend to be close to the distribution of corresponding CPU requirements. However in Dataset-3 and Dataset-4, we observe that the distribution of eligibility periods for combined requirements is not always dominated by CPU requirements. Table 4 presents the eligibility period statistics for these cases. We can observe that for the combined requirement of 2GHz+1GB, the average and median eligibility periods are closer to those for the corresponding CPU requirement than those for memory requirement. However, for combined requirement of 3GHz+2GB the median eligibility period values are closer to those for the corresponding memory requirement. It should be noted that in these two datasets, the median values of the node eligibility periods were small for 2GB requirement, whereas in the case of Dataset-1 and Dataset-2 the memory availability was high for both 1GB and 2GB requirements. From this we infer that for combined requirements with memory requirement in range of 1GB, the node eligibility is dominated by the CPU requirement. When memory requirement tends to be close to 2GB or higher, it tends to dominate the node eligibility.

To understand the distribution of eligibility set sizes we look at the statistics presented in Table 3 and the cumulative distributions given in Figure 3. We find that the eligibility set sizes decrease with the increasing capacity requirements. However, one cannot draw such a generalization for eligibility periods. We observed that for a higher capacity requirement, fewer number of nodes become eligible but some of them remain in the set for a long time. We also observe that the variation in eligibility set sizes tends to be small. This indicates that there is always some constant number of nodes that can satisfy a given requirement. For example, in case of the 4GHz CPU requirement there are always more than 18 nodes available, and for 2GHz at least 36 nodes were in the eligibility set. This is an indicator of how many tasks of a given requirement can be successfully scheduled in the system.

4 Node availability and Eligibility Period

In this section, we characterize the relationship between a node's total availability for a given resource requirement during the entire observation period and its average eligibility period for that requirement. We define *availability fraction* a of a node as the ratio of total amount of time for which the node was in the eligibility set to the total duration of the observation period T. So, if a node became eligible for n number of times and x_i denotes the i'th eligibility period then *availability fraction* a is defined as

$$a = \frac{\sum x_i}{T} \qquad (2)$$

The availability fraction and the average eligibility period of a node together indicate the 'quality' of that node. If a node has high average eligibility period and availability fraction close to 1, it indicates that the node satisfies the given resource capacity requirement for most of the time during the observation period as well as it tends to remain eligible for a long period of time. The nodes which show high availability fraction but small eligibility periods tend to enter the eligibility set multiple times but remain eligible only for short periods.

Figure 4 shows the scatter graph for availability fraction of nodes and their average eligibility period for CPU capac-

	Dataset-1 (75 hours – Nov 18-21, 2009)						Dataset-2 (97 hours – December 1-4, 2009)					
	Eligibility Period (minutes)			Unique Nodes	Eligibility Set Size		Eligibility Period (minutes)			Unique Nodes	Eligibility Set Size	
	Avg	Median	Std Dev		Avg	Std Dev	Avg	Median	Std Dev		Avg	Std Dev
1GHz CPU	315	46	472	138	103	9.29	522	145	874	134	64	15.3
2GHz CPU	103	34	221	74	51	6.24	367	50	553	92	35	6.4
3GHz CPU	218	31	376	47	38	3.15	423	356	412	54	30	4.02
4GHz CPU	163	40	284	35	25	3.72	799	359	1362	30	21	3.0
1GB Memory	650	438	494	109	105	2.7	1061	1022	578	105	84	8.4
2GB Memory	335	284	281	39	34	1.75	910	787	564	35	20	5.0
2GHz+1GB	119	48	256	50	30	4.8	392	53	552	61	20	5.06
3GHz+2GB	218	108	305	16	13	1.0	518	577	502	11	5	1.7

Table 3: Eligibility Period and Set Size Statistics under the Basic Method for Node Selection

	Dataset-3 (96 hours – Jan 22-25, 2010)		Dataset-4 (133 hours – Jan 26-Feb 1, 2010)	
	Average	Median	Average	Median
2GHz CPU	300	46	273	21
3GHz CPU	234	37	268	46
1GB Memory	864	438	1018	471
2GB Memory	253	7	339	8
2GHz CPU + 1GB Memory	214	27	245	35
3GHz CPU + 2GB Memory	131	7	113	8

Table 4: Eligibility Period Statistics for Dataset-3 and Dataset-4

ity requirement of 2GHz. From Figure 4(a), we observe that in case of Dataset-1, there are significant number of nodes which have high availability fraction but small eligibility periods. Such nodes will appear in the eligibility set for longer durations than the nodes which show small availability fraction, and hence such nodes are more likely to get selected for hosting a task. Since these nodes have small eligibility periods, selecting them for hosting a task can lead to frequent migration of the task. The profiling approach will help in eliminating such nodes from inclusion into the eligibility set. We observe that, in case of Dataset-3 and Dataset-4 number of such nodes is less.

5 Node Selection using the Profiling Method

The results of our evaluations of the profiling approach using Dataset-1 and Dataset-2 are shown in Figure 5 and Table 5. These data show a clear and remarkable benefit of using profiling. For example, comparing the eligibility period values for the 2GHz requirement for Dataset-1 using the basic method with those with the profiled method, one can notice that the average period increases from 103 to 496 minutes, and the median value also increases from 34 to 258 minutes. The CDF graphs of eligibility periods for CPU requirements for these two datasets are given in Figure 5. As expected, the eligibility set sizes are always smaller in case of the profiled approach. This means we have a smaller set of nodes in the eligibility set but they are of higher "quality", i.e. they are likely to meet the given requirement for a longer time. In the data presented in Table 5, there was only one case where the values for the eligibility period using the profiled method were smaller than those with the basic method. This occurred for 1GB memory requirement in case of Dataset-1. We have not found any clear explanation for this case. Nonetheless, there is clear evidence otherwise that the profiling method identifies better quality nodes for the eligibility sets.

5.1 Comparison of CPU and Memory Availability

In the context of conjoined requirements for CPU capacity and memory, we were interested in determining how often the requirement for one type of resource becomes the dominating factor in determining node eligibility. For this purpose we counted at each observed node the number of times the required resource capacity for one type was unavailable when the required capacity for the other type was available at that node. We conducted this evaluation for a spectrum of resource requirements as shown in Table 6. In Table 6 we present this evaluation. The data given in the column labeled "CPU" indicates the fraction of the time when the memory requirement was satisfied but the required CPU capacity was not available on a node. This indicates the fraction of the time the unavailability of the CPU capacity that prevented a node from inclusion in the eligibility set. Similarly the complementary data is presented in this table for memory unavailability.

From Table 6 we observe that, for memory requirement of 1GB or less, the node eligibility is dominated by the availability of required CPU capacity. For example, in the case of 1GHz+1GB requirement for Dataset-1, 65% of the time, a node which satisfied the memory requirement, failed to satisfy the CPU requirement. For 35% of the time a node did not have the required amount of memory but satisfied the CPU requirement. Whereas for 1GHz+2GB requirement, 16% of the time CPU capacity was not available when memory capacity was available. From this data we conclude that when memory requirement is low (around 1GB), node eligibility is determined by the CPU capacity requirement. When memory requirement is 2GB or higher, the node eligibility becomes mostly dependent on memory availability, when the CPU requirement is up to 2GHz. In the cases when both CPU and memory requirements are high, we notice that memory unavailability tends to be slightly higher.

	Dataset-1 (75 hours – Nov 18-21, 2009)						Dataset-2 (97 hours – December 1-4, 2009)					
	Eligibility Period (minutes)			Unique Nodes	Eligibility Set Size		Eligibility Period (minutes)			Unique Nodes	Eligibility Set Size	
	Avg	Median	Std Dev		Avg	Std Dev	Avg	Median	Std Dev		Avg	Std Dev
1GHz CPU	740	359	895	121	50.46	16.4	786	195	1107	99	42.8	16.49
2GHz CPU	496	258	512	39	20.67	9.42	951	216	1287	55	34.6	8
3GHz CPU	552	386	557	29	15.6	8	1312	709	1681	30	24.3	5.5
4GHz CPU	356	134	452	22	11.9	4.6	1406	1129	1567	16	12.45	2.72
1GB Memory	455	340	393	48	28.39	14.4	2203	2126	1398	90	72.9	22.8
2GB Memory	957	1241	541	20	12.47	6.31	2469	2310	1232	26	21.74	6.95
2GHz+1GB	701	711	481	21	12.33	6.52	983	347	1127	34	26.93	7.28
3GHz+2GB	885	1160	599	8	4.95	2.82	1784	1784	1790	7	5.46	1.7

Table 5: Eligibility Period and Set Size Statistics under Profiling Based Node Selection

Requirements	Dataset-1		Dataset-2		Dataset-3		Dataset-4	
(CPU+Mem)	CPU	Memory	CPU	Memory	CPU	Memory	CPU	Memory
1GHz+1GB	0.65	0.35	0.75	0.25	0.53	0.47	0.64	0.36
1GHz+2GB	0.16	0.84	0.25	0.75	0.05	0.95	0.09	0.91
2GHz+1GB	0.84	0.16	0.86	0.14	0.72	0.28	0.84	0.16
2GHz+2GB	0.35	0.65	0.35	0.65	0.12	0.88	0.20	0.80
3GHz+1GB	0.90	0.1	0.92	0.08	0.86	0.14	0.94	0.06
3GHz+2GB	0.47	0.53	0.44	0.56	0.35	0.65	0.48	0.52

Table 6: Fraction of the time the capacity for specified resource type is unavailable when the other resource type is available

6 Related Work

Several other research projects, such as CoMon [4] and Sophia [8], have investigated monitoring of PlanetLab nodes for their resource consumption. CoMon periodically collects and provides node-level statistics such as the number of active slices, per slice utilization of CPU, memory, and bandwidth. A number of research projects have analyzed this data for characterizing the resource utilization [3, 2]. The work in [2] presents statistical methods for resource discovery and for characterization of nodes based on their resource usage. It classifies nodes into different groups based on the similarities in their resource availability characteristics. The focus of the work in [3] was mainly on the characterization of resource availability of the PlanetLab nodes based on long-term observation data. In contrast, our focus is on characterization of nodes based on their recent resource availability for scheduling and dynamic relocation of migratory tasks. We also present and evaluate different approaches for selecting a node. In [7], analysis of the CoMon data is presented for characterizing node failures and availability. In contrast to these previous works, our focus is on online monitoring and selection of PlanetLab nodes for dynamic placement and relocation of tasks.

7 Conclusion

We presented our study of node availability on the PlanetLab for a spectrum of resource requirements. Our study finds that the eligibility periods can fluctuate significantly, but the eligibility set sizes tend to show relatively less variations. We have presented here two approaches for selecting nodes for a given resource requirement. One approach here takes into account the recent profile of the nodes. We have shown here that the profiling based approach tends to identify significantly better availability nodes for a given requirement. In this study we also find that in case of combined requirements, the CPU requirement tends to dominate in comparison to the memory requirement when the required memory is around 1GB or less; however, the memory tends to influence the node eligibility when memory requirement is around 2GB or higher.

References

[1] Bavier, A., Bowman, M., Chun, B., Culler, D., Karlin, S., Muir, S., Peterson, L., Roscoe, T., Spalink, T., and Wawrzoniak, M. Operating System Support for Planetary-scale Network Services. In *NSDI'04: Proceedings of the 1st conference on Symposium on Networked Systems Design and Implementation* (Berkeley, CA, USA, 2004), USENIX Association, pp. 19–19.

[2] Cardosa, M., and Chandra, A. Resource Bundles: Using Aggregation for Statistical Wide-Area Resource Discovery and Allocation. In *The 28th IEEE International Conference on Distributed Computing* (June 2008), pp. 760–768.

[3] Oppenheimer, D., Chun, B., Patterson, D., Snoeren, A. C., and Vahdat, A. Service Placement in a Shared Wide-Area Platform. In *ATEC '06: Proceedings of the Annual Conference on USENIX '06 Annual Technical Conference* (Berkeley, CA, USA, 2006), USENIX Association, pp. 26–26.

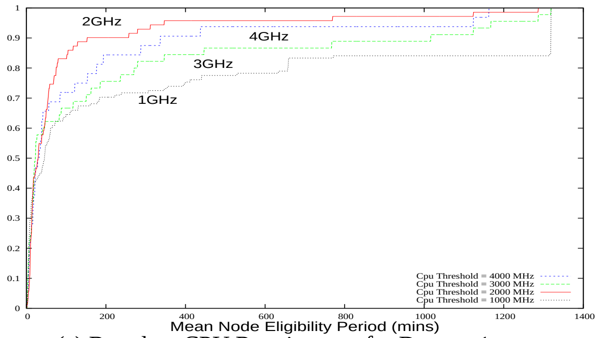
[4] Park, K., and Pai, V. S. CoMon: A Mostly-scalable Monitoring System for PlanetLab. *SIGOPS Oper. Syst. Rev.* (2006), 65–74.

[5] Tripathi, A., Karnik, N., Vora, M., Ahmed, T., and Singh, R. Mobile Agent Programming in Ajanta". In *1999 International Conference on Distributed Computing Systems* (June 1999), IEEE Computer Society.

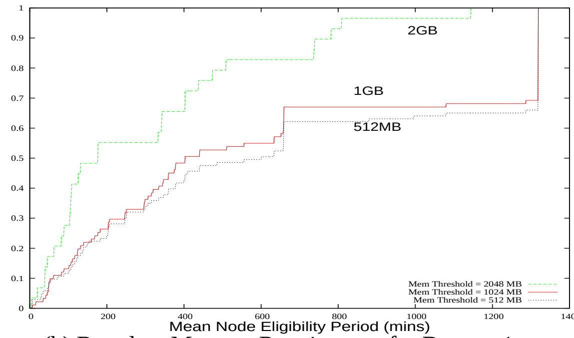
[6] Tripathi, A., Padhye, V., and Kulkarni, D. Resource-Aware Migratory Services in Wide-Area Shared Computing Environments . In *Proceedings of the IEEE Symposium on Reliable Distributed Systems (SRDS'09)* (2009), pp. 51–60.

[7] Warns, T., Storm, C., and Hasselbring, W. Availability of Globally Distributed Nodes. In *Proceedings of the IEEE Symposium on Reliable Distributed Systems* (2008), pp. 279–284.

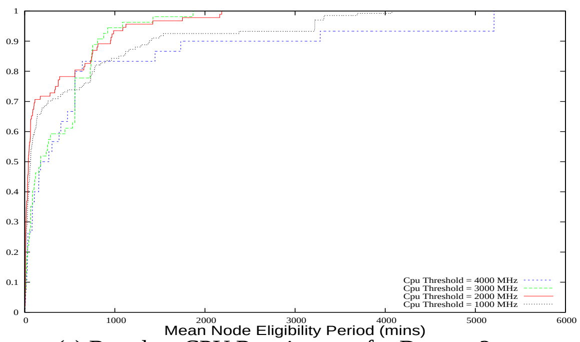
[8] Wawrzoniak, M., Peterson, L., and Roscoe, T. Sophia: An Information Plane for Networked Systems. *SIGCOMM Comput. Commun. Rev. 34*, 1 (2004), 15–20.



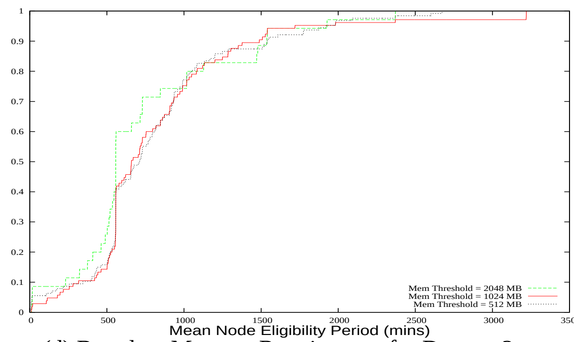
(a) Based on CPU Requirement for Dataset-1



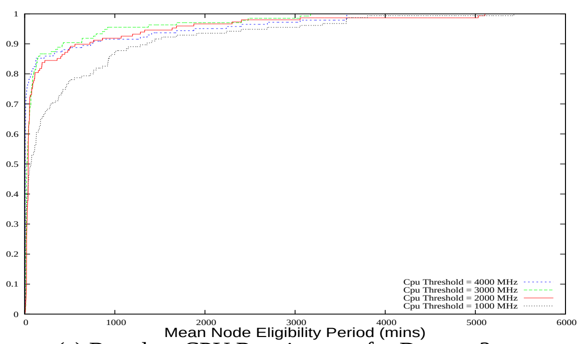
(b) Based on Memory Requirement for Dataset-1



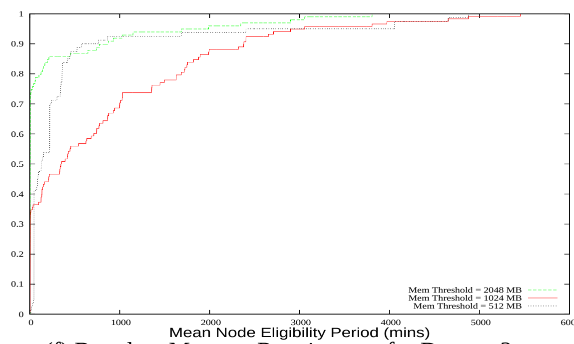
(c) Based on CPU Requirement for Dataset-2



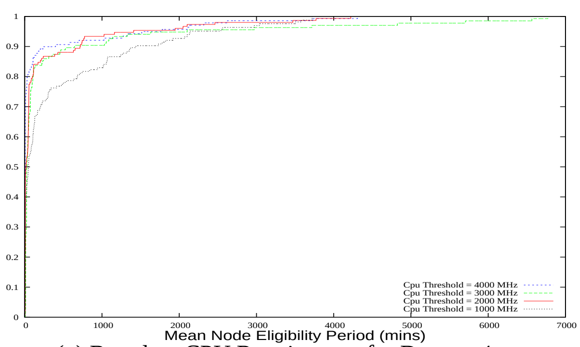
(d) Based on Memory Requirement for Dataset-2



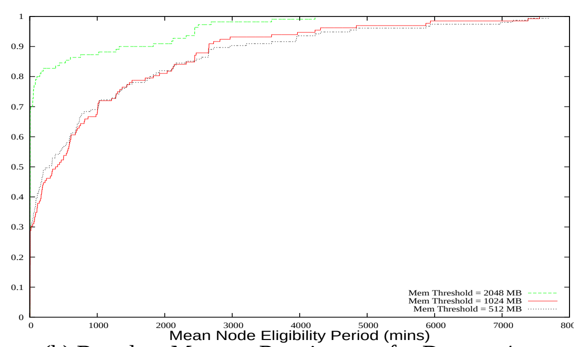
(e) Based on CPU Requirement for Dataset-3



(f) Based on Memory Requirement for Dataset-3

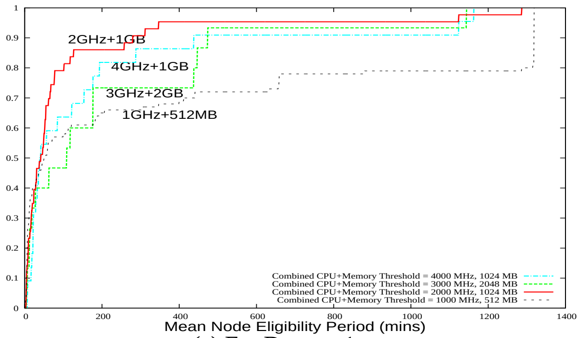


(g) Based on CPU Requirement for Dataset-4

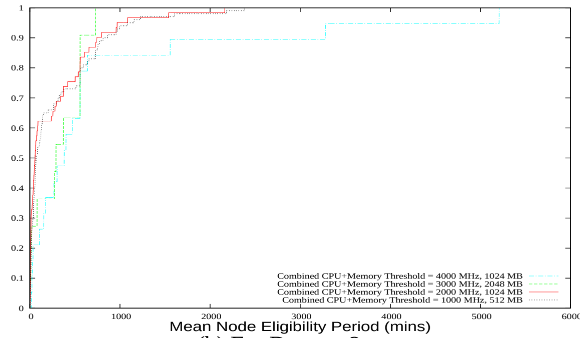


(h) Based on Memory Requirement for Dataset-4

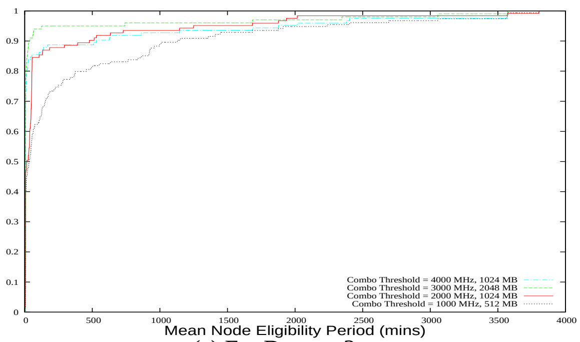
Figure 1: Cumulative distributions of Eligibility Periods for CPU and Memory Requirements



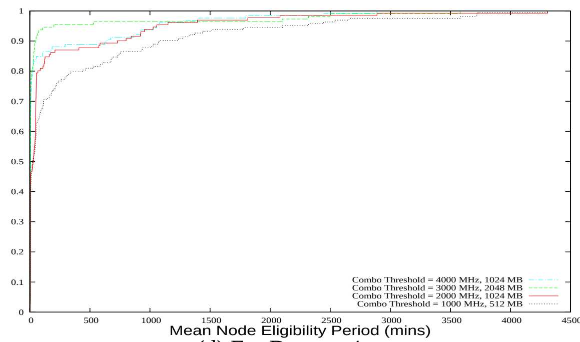
(a) For Dataset-1



(b) For Dataset-2

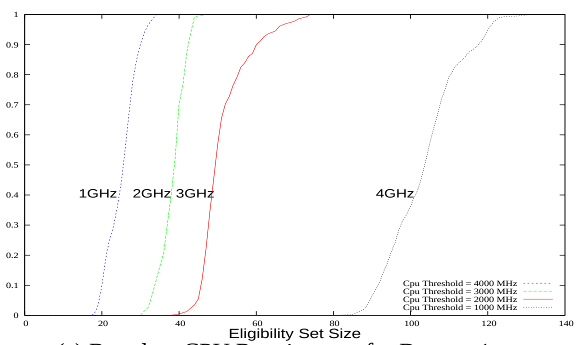


(c) For Dataset-3

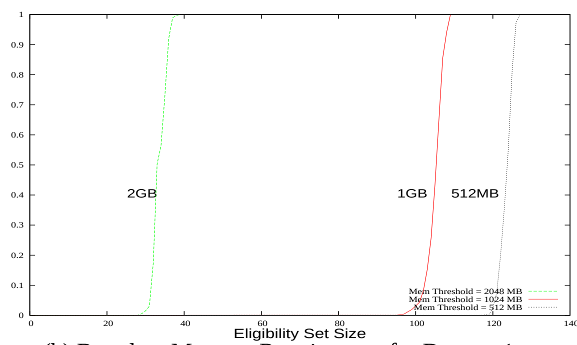


(d) For Dataset-4

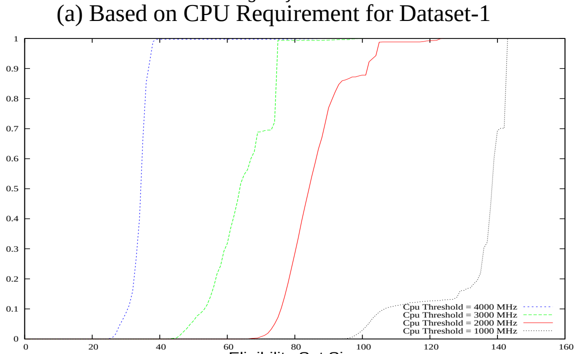
Figure 2: Cumulative distributions of Eligibility Periods for Combined Requirements



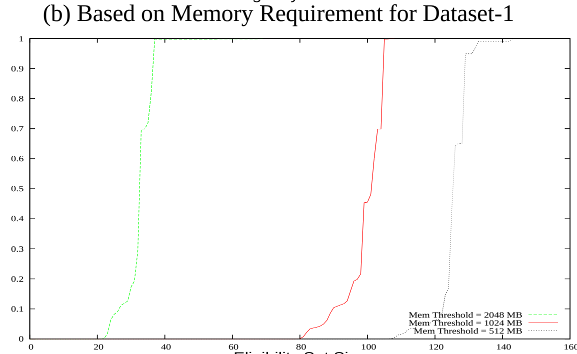
(a) Based on CPU Requirement for Dataset-1



(b) Based on Memory Requirement for Dataset-1

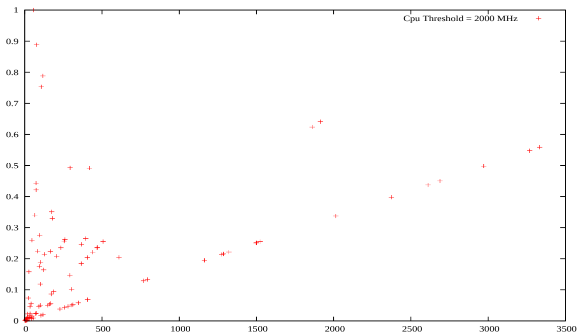


(c) Based on CPU Requirement for Dataset-3

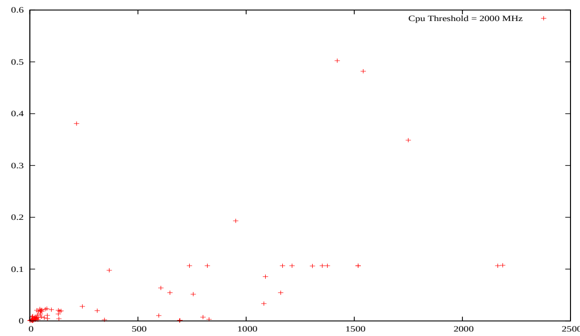


(d) Based on Memory Requirement for Dataset-3

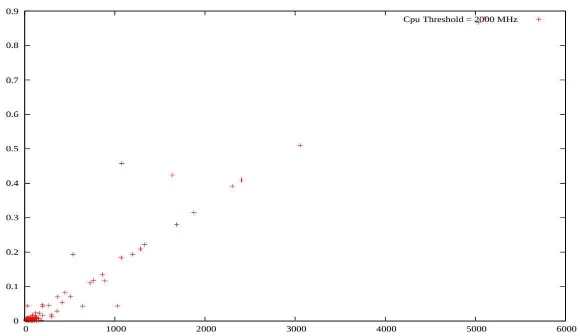
Figure 3: Cumulative distributions of Eligibility Set Size for CPU and Memory Requirements



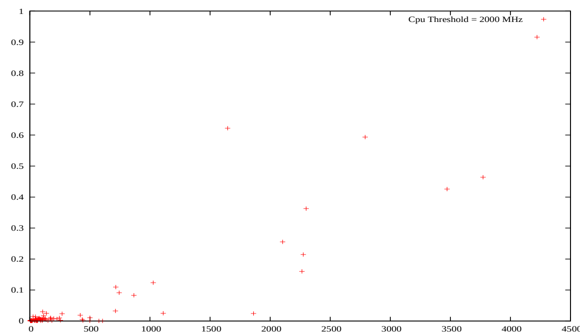
(a) Based on Dataset-1



(b) Based on Dataset-2

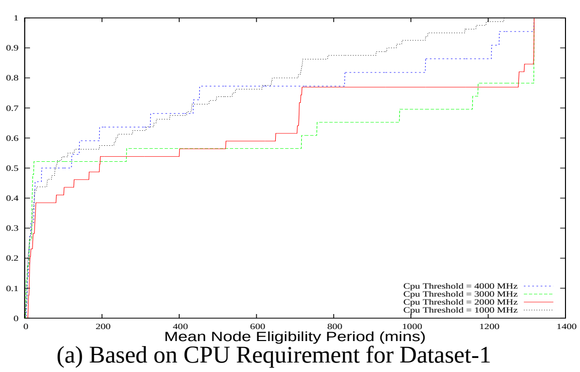


(c) Based on Dataset-3

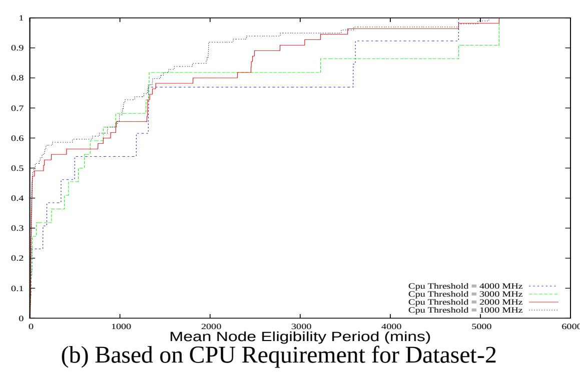


(d) Based on Dataset-4

Figure 4: Availability fraction vs. Average Eligibility Periods for eligible nodes (for CPU requirement of 2 GHz)



(a) Based on CPU Requirement for Dataset-1



(b) Based on CPU Requirement for Dataset-2

Figure 5: Cumulative Distribution of Eligibility Periods Based on Profiling (for CPU requirements)