CHAPTER IV

EXPERIMENTAL RESULTS AND DISCUSSION

This chapter describes the results and discusses the experiments. First, experimental data are described, followed by experimented methods. The chapter concludes with method comparison and discussion of the research results.

4.1 Data set of the experiment

Data used in this research came from Suan Sunandha Rajabhat University. The data were separated into two parts. First, user behavior consisted of user access behavior and data size (in byte) required by the users from proxy servers and web servers during a 60 day period. Second, data concerning the use of CPU and memory were collected from the three servers, i.e., database server, web server, and application server for every 10 seconds during the period of 60 days.

In the experiment, data for 45 days were used for creating model and the remaining data were for testing.

4.2 Experiment 1: Results of User Behavior Analysis

Performance of the model was tested. In general, the data was divided into a training data set and a test data set. Data obtained in November and December for 45 days were used to train the model, while data acquired for 15 days in December were used to test the performance of the model. Note that the ratio of the training set and testing set is 75:25.

In Table 4.1, the prediction model uses association rules with training set on Monday 12.00 AM – 8:00 AM, having the level of user access for proxy server "Low." The rule is verified by test set using the same day and time. The level of the user access is "Low", which corresponds to the generated association rule. Therefore, the accuracy is marked as "T" (Test result is accurate).

Table 4.1. Example of test model on Monday for proxy server

Rule (Training Set)	Test Set	Accuracy
Monday ,12:00 AM =>Low	Low	T
Monday,01:00 AM =>Low	Low	T
Monday, 02:00 AM =>Low	Low	T
Monday,03:00 AM =>Low	Low	T
Monday,04:00 AM =>Low	Low	T
Monday,05:00 AM =>Low	Low	T
Monday,06:00 AM =>Low	Low	T
Monday,07:00 AM =>Low	Low	T
Monday,08:00 AM =>Low	Low	T
Monday,09:00 AM =>Medium Low	Medium Low	T
Monday,10:00 AM =>Low	Medium High	F
Monday,11:00 AM=>Medium Low	Medium	F
Monday,12:00 PM =>Medium Low	Medium	F
Monday,13:00 PM =>Low	Medium	F
Monday,14:00 PM=>Medium Low	Low	F
Monday,15:00 PM=>Medium Low	Medium Low	T
Monday,16:00 PM =>Low	Low	T
Monday,17:00 PM =>Low	Low	T
Monday,18:00 PM=>Low	Low	T
Monday,19:00 PM =>Low	Low	T
Monday,20:00 PM =>Low	Low	T
Monday,21:00 PM =>Low	Low	T
Monday,22:00 PM =>Low	Low	T
Monday,23:00 PM =>Low	Low	T

However, on Monday 10.00 AM., the prediction model using association rules indicates a "Low" level of user access. The rule is validated by the test set using the same day and time. The level of user access is "Medium High", which does not correspond to the rules. Therefore, the accuracy is marked as "F" (Test result is inaccurate).

Performance of the predictive model for the proxy server is 86.86%. Similarly, performance of the predictive model for the web server is measured in the same manner as the proxy server, yielding 87.18%.

4.3 Experiment 2: Results of Hardware Resources Consumption Analysis

Since the proposed approach concerns two related issues, i.e., predicting the requested resources and resource allocation to satisfy the response time, the experimental results relevant to each issue will be separately addressed as follows.

The percentage of CPU and memory usages in each hour predicted by simple exponential and double exponential smoothing methods were compared and shown in Figure 4.1. The predicted CPU usages of database, application, and web servers are shown in Figure 4(a), (c), and (e), respectively. Predicted memory usages of those three different servers are shown in Figure 4(b), (d), and (f), respectively. Notice that the predicted CPU and memory usage of all servers by double exponential smoothing method are more accurate than those from simple exponential smoothing method.

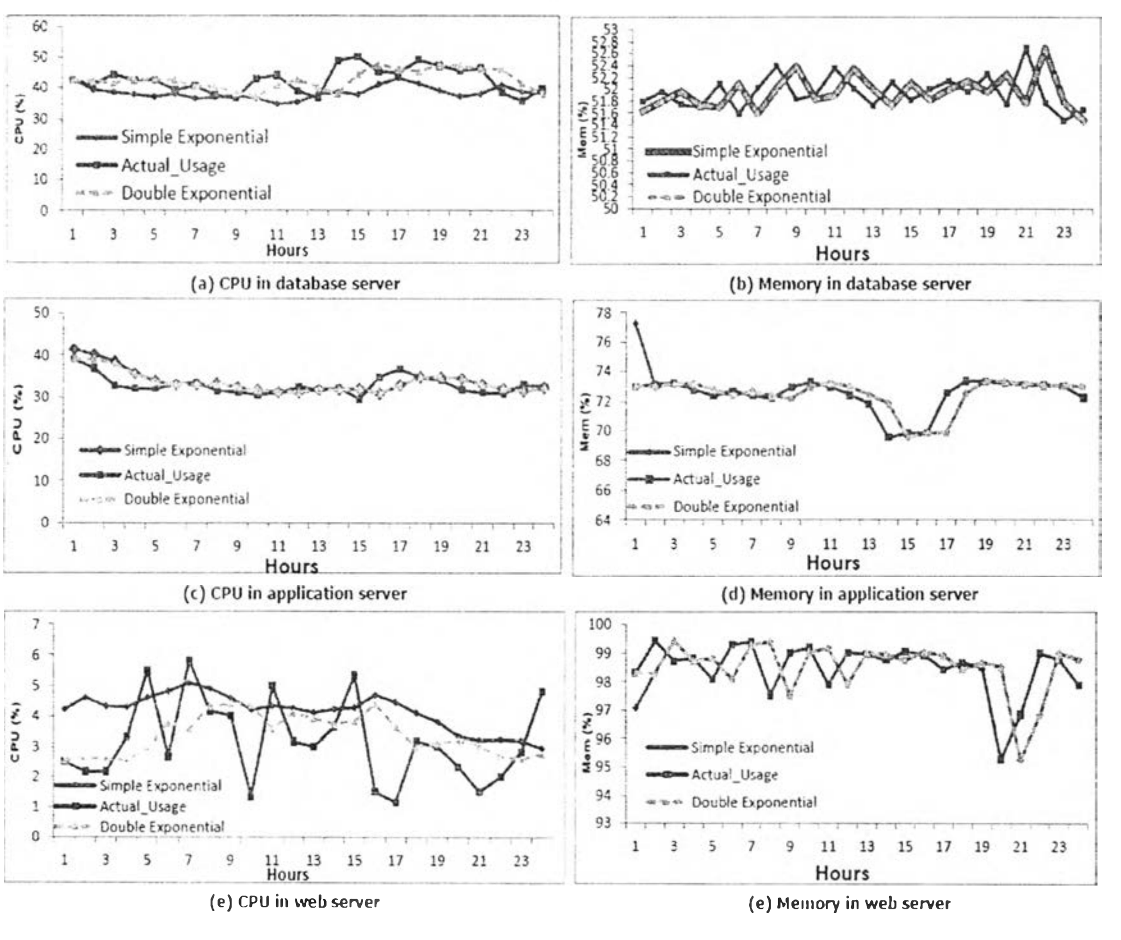


Fig.4.1. Resource prediction by exponential smoothing method for three servers.

Table 4.2: Sum square error with alpha values for simple exponential.

Alpha	SSE
0.6	11449.31
0.7	11511.47
0.5	11517.84
0.8	11692.58
0.4	11751.99
0.9	11994.12
0.3	12240.2
1	12427.98
0.2	13242.28
0.1	15824.27

Table 4.3: Sum square error with alpha and gamma values for double exponential.

Alpha	Gamma	SSE
0.6	0	11413.69
0.5	0	11478.14
0.7	0	11478.18
0.8	0	11660.44
0.4	0	11705.23
0.9	0	11962.21
0.3	0	12180.82
1	0	12395.51
0.2	0	13158.82
0.6	0.2	13324.97

Table 4.2 shows the values of alpha are the corresponding sum square error (SSE) in simple exponential technique. The model has the best prediction when assign alpha = 0.6. Likewise, Table 4.3 shows the values of alpha and gamma in double exponential are the corresponding sum square error from model prediction. Alpha was set to 0.6 as the same simple exponential and gamma was set to zero, the model gave the minimum error.

4.4 Experiment 3: Comparison with Other Predicting Models

Two predicting methods, i.e., association rules and ARIMA, were compared with the propose model in term of prediction accuracy.

4.4.1 Comparing with Association Rule

In this section, the items of the left-hand side are the day, minute, and second, while the items on the right-hand side are the amount of allocated resources in advance which is defined in term of level unit. Some rules are eliminated by using *sup* and *conf* measures.

Resource usages predicted by association rules in this experiment were classified into the following five levels: *1* for low level; *2* for medium low level; *3* for medium level; *4* for medium high level; and *5* for high level. Each level is determined by a set of rules of the form (D, T)⟶ L, where D is the day, T_j is the time when the usage occurs, and L is the level. For example, the rule shown below means resource usage on Monday at 06:00PM is in level 3.

(Monday, 06:00 PM) ⟶ 3

Table 4.4 shows examples of rules for predicting memory usage levels on Monday and Tuesday at 06:00 PM in the database server. Confidence and support values are used for rule selections. It may be possible that there is more than one established rule with equal confidence. In this situation, the rule with maximum support will be selected. But if there exists a set of rules whose confidence and support values are equal, then the first rule of this set will be selected [41].

Table 4.5 shows examples association rule prediction of memory usage by the database server during 24x7 hours of service.

Table 4.4. Examples of association rules for predicting.

rule	Conf (%)	Sup (%)
Monday , 06:00 PM ⟶ 3	37.5	0.219
Monday , 06:00 PM ⟶ 4	50	0.292
Monday , 06:00 PM ⟶ 2	12.5	0.07
Monday , 06:00 PM ⟶ 4	50	0.292
Tuesday , 06:00 PM ⟶ 1	12.5	0.073
Tuesday, 06:00 PM ⟶ 3	25	0.146
Tuesday, 06:00 PM ⟶ 4	50	0.292
Tuesday, 06:00 PM ⟶ 4	50	0.292

Table 4.5. Examples of predicting by association rule with confidence and support.

No.	rule	Conf (%)	Sup (%)
1	Monday , 0:00 ⟶ 3	62.5	0.365
...	....	...	...
24	Monday , 23:00 ⟶ 3	62.5	0.365
25	Tuesday , 0:00 ⟶ 3	50	0.292
...	....	...	...
168	Sunday , 23:00 ⟶ 3	55.56	0.365

From association rule with confidence and support values in Table 4.5, the levels of memory usage predicted for database server in each hour period of 7 days are summarized in Table 4.6. To compare this predicting method with other methods, each level number must be converted to a numeric value corresponding to the resource usage as follows. At level *i*, there may be different numbers of resource usage.

Let $R^{(i)} = \{r_1^{(i)},..., r_n^{(i)}\}$ be the set of different resource usage $r_j^{(i)}$ at level *i*. The numeric value of resource usage at level *i* is equal to $\sum_{j=1}^{n} \frac{r_j^{(i)}}{|R^{(i)}|}$

Table 4.6. Level of memory usage in 24 hours by association rules.

Time	sat	sun	mon	tue	wed	thu	fri
0	4	2	3	3	2	4	2
1	4	2	3	3	3	2	3
2	3	2	3	3	3	4	4
3	3	2	3	3	3	4	3
4	4	2	4	3	2	5	2
5	4	2	4	2	3	1	2
6	2	2	3	3	3	4	2
7	5	2	4	3	3	4	2
8	4	2	3	3	2	4	2
9	3	2	3	3	3	2	2
10	4	2	4	3	3	3	3
11	4	2	2	3	3	2	3
12	2	2	3	3	4	3	3
13	2	2	3	3	4	2	4
14	2	3	3	3	4	2	4
15	4	3	3	3	4	3	4
16	2	3	3	3	4	4	3
17	2	2	3	4	4	3	3
18	2	3	4	4	4	2	4
19	3	3	2	3	4	2	4
20	3	3	3	3	4	3	3
21	3	2	3	3	4	3	4
22	3	3	3	2	4	3	4
23	2	3	3	2	4	3	2

Table 4.7.Memory usage in percentage of 24 hours by association rules.

Time	sat	sun	mon	tue	wed	thu	fri
0	52	40	46	46	40	52	40
1	52	40	46	46	46	40	46
2	46	40	46	46	46	52	52
3	46	40	46	46	46	52	46
4	52	40	52	46	40	100	40
5	52	40	52	40	46	34	40
6	40	40	46	46	46	52	40
7	100	40	52	46	46	52	40
8	52	40	46	46	40	52	40
9	46	40	46	46	46	40	40
10	52	40	52	46	46	46	46
11	52	40	40	46	46	40	46
12	40	40	46	46	52	46	46
13	40	40	46	46	52	40	52
14	40	46	46	46	52	40	52
15	52	46	46	46	52	46	52
16	40	46	46	46	52	52	46
17	40	40	46	52	52	46	46
18	40	46	52	52	52	40	52
19	46	46	40	46	52	40	52
20	46	46	46	46	52	46	46
21	46	40	46	46	52	46	52
22	46	46	46	40	52	46	52
23	40	46	46	40	52	46	40

Table 4.7 shows the numeric value of corresponding memory usage after converting from its level. The average percentage of resource usage for each hour is converted from the level usage.

Fig. 4.2 shows the comparison of prediction results obtained from double exponential smoothing, association rules, and actual resource usage for three different servers. The results of the proposed method are more accurate than others for the three servers.

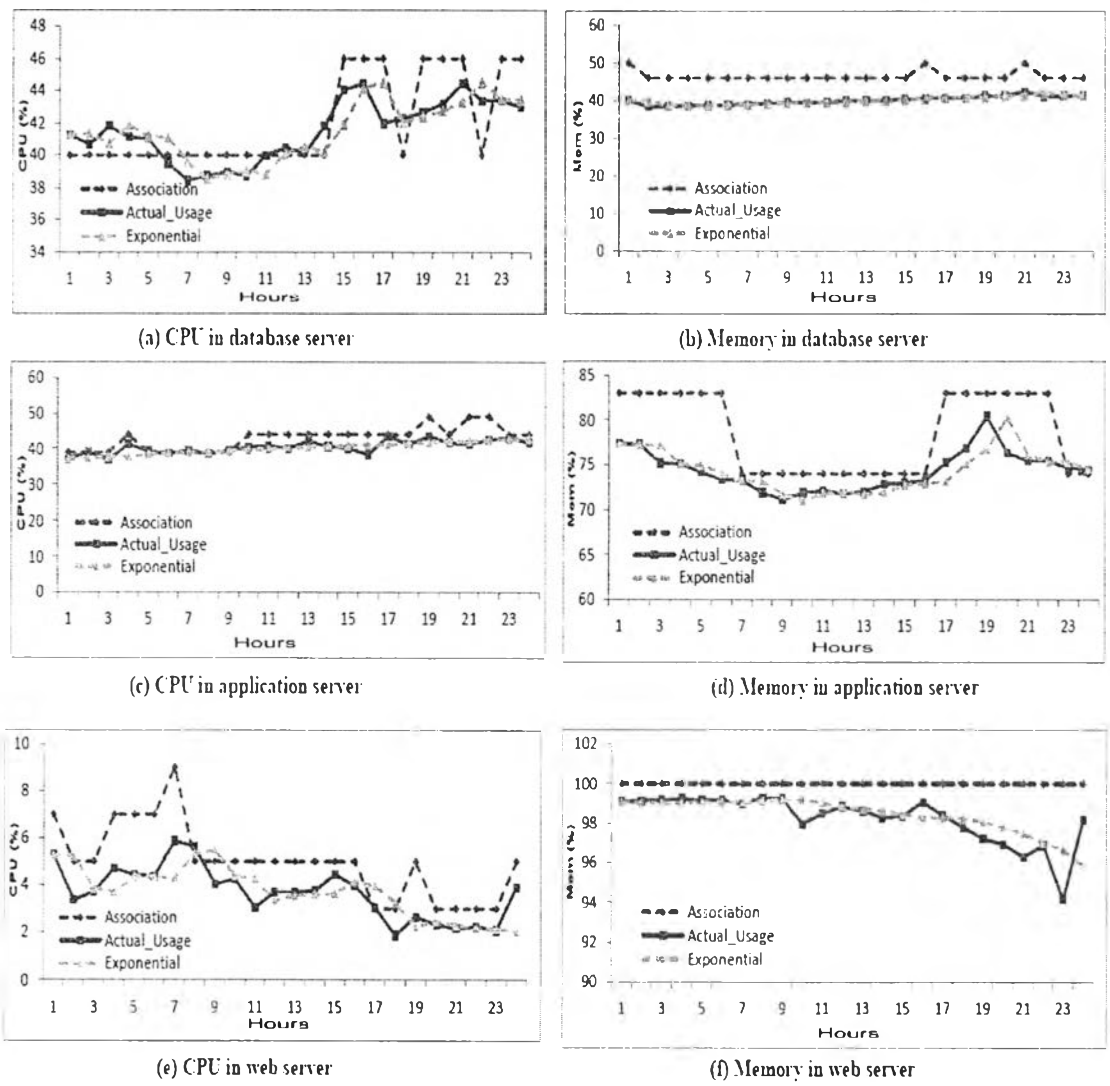
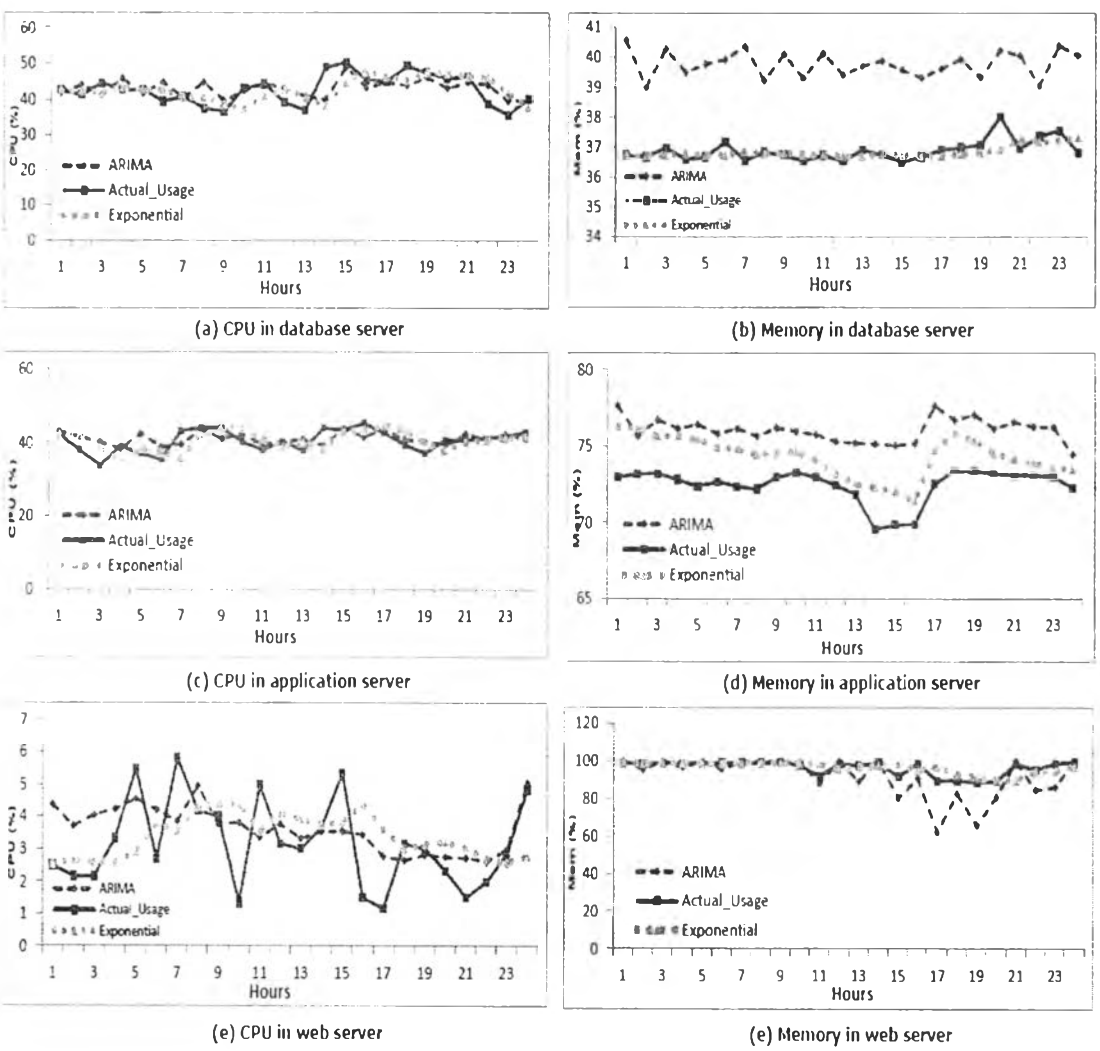


Fig.4.2. Comparison of resource prediction by association rules and the proposed method using double exponential smoothing method for three servers.

4.4.2 Comparing with ARIMA

ARIMA is an efficient statistical forecasting method. The results from the proposed method were compared with ARIMA for CPU and memory usage in three servers. Fig. 4.3 shows the comparison results obtained from double exponential smoothing, ARIMA, and actual resource usage. Generally, the proposed method produces higher accuracy than ARIMA except for the case of CPU usage of web server. In this case, the errors from both methods were almost the same which were rather high. The best result was memory usage prediction of database and web servers.



(a) CPU in database server

(b) Memory in database server

(c) CPU in application server

(d) Memory in application server

(e) CPU in web server

(e) Memory in web server

Fig.4.3. ARIMA model resource prediction of three servers.

4.4.3 Summary of Comparison

Measure of mean square error (MSE) was used to evaluate the performance of each method with respect to the actual values. MSE is defined by the following equation. Suppose there are *n* data points.

$$MSE=\sum_{i=1}^{n} (y_i-\hat{y}_i)^2/n \qquad (4.1)$$

Table 4.8. MSE Comparison for method prediction

Method Prediction	Database		Application		Web	
	CPU	Memory	CPU	Memory	CPU	Memory
Association	62.06	163.03	13.65	153.24	2.06	28.12
ARIMA	13.43	3.87	8.26	14.71	1.81	13.33
Simple Exponential	8.64	0.17	5.41	1.43	1.59	1.38
Double Exponential	7.16	0.17	4.52	0.66	1.59	1.31

where Y_i is the actual value at time *i*. $\hat{y}_i$ is the predicted value at time *i*. Table 4.8 shows MSE for all method predictions in three servers. Association rules provide the highest prediction error. This is probably because resource usage and the period of time in each day are less correlated. Furthermore, ARIMA model prediction is more efficient than association rules. Simple and Double exponential techniques give slightly different values. Double exponential smoothing technique gives the least MSE. Due to data from three servers having different functions, resource usage of CPU and memory units also varied. Double exponential can predict data whose behaviors are different from stationary or non-stationary because this model adjusts the smoothing constant between actual data and predicting value (α) and the smoothing constant between the trend of actual data and the trend of prediction value (γ).

4.5 Experiment 4: Effect of Compromising Factor on Resource Allocation

Generally, the predicted number of resources may be higher or lower than the actual number of requested resources. If the predicted number of resources is higher than the actual requested one, the best response time can be achieved. But if the predicted number of resources is lower than the requested one, additional resources must be augmented to the predicted one to maintain the best response time. After the prediction by exponential smoothing method, more number of CPUs and memory units are augmented. However, the augmented resources must not be too many to affect optimizing the number of idle resources and power consumption of those idle resources. Determining the optimal number of augmented resources is not straightforward since the actual behavior of CPU and memory requests are unknown in advance. The only known information is the predicted request. In the experiment, an additional compromising factor u is empirically set to 65%-75% of resource utilization as optimize resource usage period.

Fig. 4.4 shows CPU allocation, actual CPU usage, and the predicted values. Fig. 4.5 shows memory allocation, memory usage, and the predicted values. The value of compromising factor u can directly affect the performance if utilization is close to the knee value. On the other hand, if utilization is set too low, it would be wasteful since there may be too many idle resources.

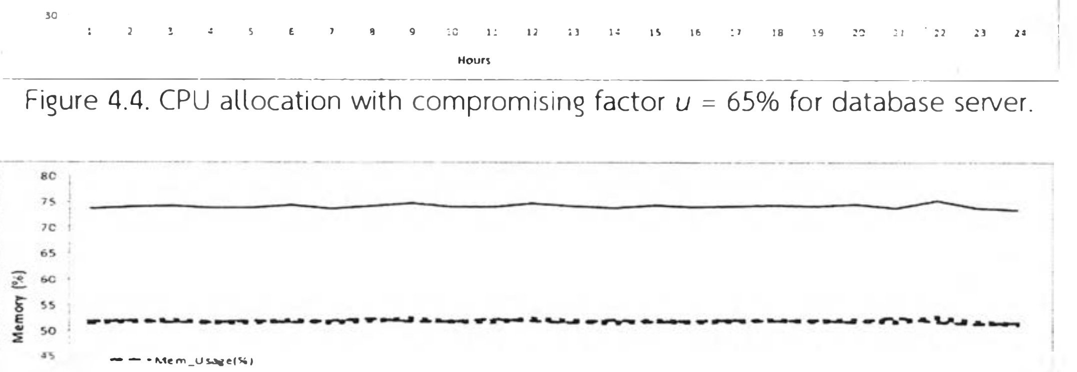


Figure 4.4. CPU allocation with compromising factor u = 65% for database server.

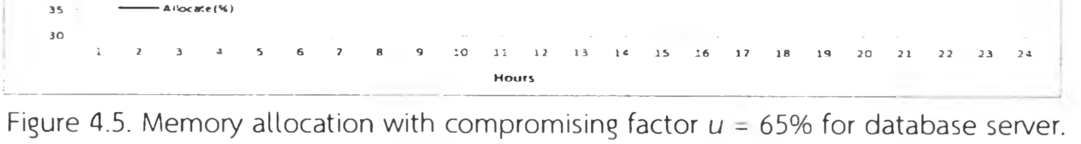


Figure 4.5. Memory allocation with compromising factor u = 65% for database server.

But higher utilization will deteriorate response time because the allocated resources may not be enough to run the given task, which implies that all allocated resources must be busy all the time. Example in Table 4.9, at 2 o'clock, the predicted value is 37.17. If the value of *u* is set to 65% (*u*65), the allocated value becomes (37.17*100)/65 = 57.19. But the actual resource usage is equal to 38. Hence, the real utilization is (38*100)/57.19 = 66.45. This value is over the utilization boundary but not over 75%.

In contrast to previous 65% utilization, at 2 o'clock if the value of *u* is set to 75% (*u*75) of the predicted value of 37.17, the allocated resource becomes (37.17*100)/75 = 49.56. But the actual resource usage is 38. Hence, the real utilization is (38*100)/49.56 = 76.67, which exceeds utilization boundary of 75%.

Table 4.10 demonstrates the utilization for different values of *u*, denoted by *u*65, *u*66, *u*67, *u*68, *u*69, *u*70, and *u*75. The results indicate that when the value of *u* increases, utilization of the system may also increase in some cases. This may reduce response time of the system.

Table 4.9. Resource utilization with allocation at *u*65 and *u*75.

Time	Predicted (%)	CPU_Usage	Allocated With *U65*	Utilization (65%)	Allocated With *U75*	Utilization (75%)
1	37.17	37.17	57.18	65	49.56	75
2	37.17	38	57.19	66.45	49.56	76.67
3	37.67	37.5	57.96	64.7	50.23	74.66
4	37.57	36.67	57.8	63.44	50.1	73.2
5	37.03	38.67	56.97	67.87	49.38	78.31
6	38.02	35.67	58.49	60.99	50.69	70.37
7	36.61	37.17	56.32	65.99	48.81	76.15
8	36.95	38	56.84	66.85	49.26	77.14
9	37.58	36	57.82	62.26	50.11	71.84
10	36.64	33.5	56.36	59.43	48.85	68.58
11	34.76	35.83	53.47	67.01	46.34	77.31
12	35.41	39.5	54.47	72.51	47.21	83.67
13	37.87	39.33	58.26	67.51	50.49	77.9
14	38.75	37.33	59.62	62.62	51.67	72.25
15	37.9	43.17	58.31	74.03	50.54	85.42
16	41.06	44.83	63.18	70.96	54.75	81.88
17	43.33	40.17	66.66	60.26	57.77	69.53
18	41.43	37.67	63.75	59.09	55.25	68.19

Time	Predicted (%)	CPU_Usage	Allocated With *U65*	Utilization (65%)	Allocated With *U75*	Utilization (75%)
19	39.18	36	60.27	59.73	52.24	68.92
20	37.27	38.67	57.35	67.43	49.7	77.81
21	38.11	42.83	58.64	73.04	50.82	84.28
22	40.95	37.33	63	59.26	54.6	68.37
23	38.78	38	59.67	63.69	51.71	73.48
24	38.32	34.83	58.95	59.08	51.09	68.17

Table 4.10. Resource utilization with different utilization boundary.

Time	*U65*	*U66*	*U67*	*U68*	*U69*	*U70*	*U75*
1	65.00	66.00	67.00	68.00	69.00	70.00	75.00
2	66.45	67.47	68.49	69.51	70.54	71.56	76.67
3	64.70	65.70	66.69	67.69	68.68	69.68	74.66
4	63.44	64.41	65.39	66.37	67.34	68.32	73.20
5	67.87	68.92	69.96	71.01	72.05	73.09	78.31
6	60.99	61.93	62.86	63.80	64.74	65.68	70.37
7	65.99	67.01	68.03	69.04	70.06	71.07	76.15
8	66.85	67.88	68.91	69.94	70.96	71.99	77.14
9	62.26	63.22	64.18	65.14	66.09	67.05	71.84
10	59.43	60.35	61.26	62.18	63.09	64.01	68.58
11	67.01	68.04	69.07	70.10	71.13	72.16	77.31
12	72.51	73.63	74.75	75.86	76.98	78.09	83.67
13	67.51	68.55	69.59	70.63	71.67	72.70	77.90
14	62.62	63.58	64.54	65.51	66.47	67.43	72.25
15	74.03	75.17	76.31	77.45	78.59	79.73	85.42
16	70.96	72.05	73.14	74.24	75.33	76.42	81.88
17	60.26	61.19	62.11	63.04	63.97	64.90	69.53
18	59.09	60.00	60.91	61.82	62.73	63.64	68.19
19	59.73	60.65	61.57	62.49	63.40	64.32	68.92
20	67.43	68.47	69.51	70.54	71.58	72.62	77.81
21	73.04	74.17	75.29	76.41	77.54	78.66	84.28
22	59.26	60.17	61.08	61.99	62.90	63.81	68.37
23	63.69	64.67	65.65	66.63	67.61	68.59	73.48
24	59.08	59.99	60.90	61.81	62.72	63.63	68.17