Supplementary Material for Optimal Task Dispatching on Multiple Heterogeneous Multiserver Systems with Dynamic Speed and Power Management

Keqin Li, *Fellow, IEEE*

1 Minimizing Average Cost-Performance Ratio

For a multiserver system S_i, our performance measure is $1/T_i$, which is inversely proportional to the average task response time T_i, the higher, the better. There are many different factors which determine the cost of cloud computing. Since the number of servers m_i is fixed in dynamic speed and power management, our cost measure is mainly the cost of power consumption P_i, the lower, the better. The cost-performance (or price-performance) ratio (CPR) refers to a product's ability to deliver performance for its cost. Generally speaking, products with a lower CPR are more desirable, excluding other factors. In this paper, we define CPR as cost/performance, i.e., $R_i = P_i T_i$. The average cost-performance ratio R of a group of n heterogeneous multiserver systems S_1, S_2, ..., S_n is

$$\begin{aligned} R(\lambda_1, \lambda_2, ..., \lambda_n) &= \frac{\lambda_1}{\lambda} R_1 + \frac{\lambda_2}{\lambda} R_2 + \cdots + \frac{\lambda_n}{\lambda} R_n \\ &= \frac{\lambda_1}{\lambda} P_1 T_1 + \frac{\lambda_2}{\lambda} P_2 T_2 + \cdots + \frac{\lambda_n}{\lambda} P_n T_n, \end{aligned}$$

where R is treated as a function of load distribution $\lambda_1, \lambda_2, ..., \lambda_n$.

In this section, we formulate and solve our optimal task dispatching problem with minimized average cost-performance ratio for multiple heterogeneous multiserver systems with dynamic d-speed and power management.

1.1 Problem Definition

Our *optimal task dispatching problem with minimized average cost-performance ratio* for multiple heterogeneous multiserver systems with dynamic d-speed and power management can be specified as follows: given the number n of multiserver systems, the sizes of the multiserver systems $m_1, m_2, ..., m_n$, a d_i-speed scheme $\psi_i = (b_{i,1}, b_{i,2}, ..., b_{i,d_i-1}, s_{i,1}, s_{i,2}, ..., s_{i,d_i})$ of S_i, for all $1 \le i \le n$, the power consumption model parameters $\xi_1, \alpha_1, \xi_2, \alpha_2, ..., \xi_n, \alpha_n$, the base power consumption $P_1^*, P_2^*, ..., P_n^*$, the average task execution requirement $\bar{r}$, and the task arrival rate λ, find a load distribution, i.e., the task arrival rates $\lambda_1, \lambda_2, ..., \lambda_n$ to the multiserver systems, such that the average cost-performance ratio $R(\lambda_1, \lambda_2, ..., \lambda_n)$ is minimized, subject to the constraint

$$F(\lambda_1, \lambda_2, ..., \lambda_n) = \lambda,$$

where

$$F(\lambda_1, \lambda_2, ..., \lambda_n) = \lambda_1 + \lambda_2 + \cdots + \lambda_n,$$

and $\rho_i < 1$, for all $1 \le i \le n$.

1.2 An Algorithm

The above optimization problem can be solved by using the method of Lagrange multiplier, i.e.,

$$\nabla R(\lambda_1, \lambda_2, ..., \lambda_n) = \phi \nabla F(\lambda_1, \lambda_2, ..., \lambda_n),$$

that is,

$$\frac{\partial R}{\partial \lambda_i} = \phi \frac{\partial F}{\partial \lambda_i} = \phi,$$

for all $1 \le i \le n$, where ϕ is a Lagrange multiplier.

In the following, we give $\partial R / \partial \lambda_i$. It is clear that

$$\frac{\partial R}{\partial \lambda_i} = \frac{1}{\lambda} \left(P_i T_i + \lambda_i T_i \frac{\partial P_i}{\partial \lambda_i} + \lambda_i P_i \frac{\partial T_i}{\partial \lambda_i} \right),$$

where $\partial T_i / \partial \lambda_i$ is derived in Section 6.2, and $\partial P_i / \partial \lambda_i$ is derived in Section 7.2, for all $1 \le i \le n$.

1.3 Numerical Data

Let us again consider the same group of heterogeneous multiserver systems specified in Section 6.3. In Tables 1 and 2, for the idle-speed model and the constant-speed model respectively, we show the optimal load distribution $\lambda_1, \lambda_2, ..., \lambda_7$ which gives the minimized cost-performance ratio for $\lambda = (2j-1)\lambda_{\text{step}}$, where $\lambda_{\text{step}} = \lambda_{\max}/20$ and $j = 1, 2, 3, ..., 10$. We again observe the situations of underflow. Let us consider

$$\gamma_i = P_i T_i + \lambda_i T_i \frac{\partial P_i}{\partial \lambda_i} + \lambda_i P_i \frac{\partial T_i}{\partial \lambda_i},$$

where $1 \le i \le n$. It is required that $\gamma_i = \lambda \phi$ for all $1 \le i \le n$. It is clear that $\gamma_i \ge P_i T_i$, where $T_i \ge \bar{r}/s_{i,1}$, and $P_i \ge m_i P_i^*$ for

• K. Li is with the Department of Computer Science, State University of New York, New Paltz, New York 12561, USA.
E-mail: lik@newpaltz.edu

TABLE 1
Example of Optimal Load Distribution for Minimized Cost-Performance Ratio (Idle-Speed Model)

λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
2.6100000	0.9574761	0.9415762	0.2369836	0.2369821	0.2369821	0.0000000	0.0000000
7.8300000	1.4467344	1.3985414	1.3820690	1.3681445	1.3662246	0.4341431	0.4341431
13.0500000	1.7571112	1.6838818	2.1994517	2.1325984	2.1089169	1.5840442	1.5839958
18.2700000	2.0643234	1.9652721	2.9123505	2.7779269	2.7021726	2.9289330	2.9190214
23.4900000	2.4361946	2.3094640	3.6685186	3.4568634	3.2927648	4.2104035	4.1157912
28.7100000	2.8755181	2.7437693	4.4701536	4.2275181	3.9348411	5.3616528	5.0965470
33.9300000	3.3395905	2.9999938	5.2396723	4.9999804	4.7337467	6.5224936	6.0945225
39.1500000	3.6568482	3.7425946	5.7341074	6.0440143	5.6137959	7.3329684	7.0256713
44.3700000	3.9766786	3.8999954	6.2185491	6.4999963	7.3050355	8.0760285	8.3937166
49.5900000	4.2818911	4.5543238	6.6750013	7.5870989	8.4742251	8.6911228	9.3263370

TABLE 2
Example of Optimal Load Distribution for Minimized Cost-Performance Ratio (Constant-Speed Model)

λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
2.6100000	1.3408770	1.2691230	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
7.8300000	1.7479326	1.6497842	1.5615022	1.4425676	1.4282134	0.0000000	0.0000000
13.0500000	2.1406951	2.0243656	3.0839431	2.8903756	2.7653711	0.0728928	0.0728925
18.2700000	2.1693242	2.0518267	3.1499107	2.9537938	2.8215272	2.5708045	2.5528129
23.4900000	2.4324395	2.3056368	3.6968282	3.4784685	3.2813371	4.2152945	4.0799954
28.7100000	2.8511991	2.7264940	4.4533683	4.2315910	3.9298124	5.4068518	5.1106835
33.9300000	3.3197387	2.9999938	5.2181232	4.9999804	4.7344858	6.5377437	6.1199344
39.1500000	3.6501264	3.7341058	5.7275855	6.0334231	5.6115765	7.3400449	7.0531378
44.3700000	3.9763021	3.8999954	6.2188579	6.4999963	7.2993685	8.0790354	8.3964443
49.5900000	4.2818938	4.5543029	6.6750672	7.5870616	8.4740693	8.6913027	9.3263024

the idle-speed model and $P_i \geq m_i(\xi_i s_{i,1}^{\alpha_i} + P_i^*)$ for the constant-speed model. Hence, if λ is too small, the condition $\gamma_i = \lambda\phi$ may not be satisfied by some multiserver system S_i. In this case, we have to set $\lambda_i = 0$. For instances, in Table 1, the above situation happens to S_6 and S_7 when $\lambda = 2.61$. In Table 2, the above situation happens to S_3, S_4, S_5, S_6, S_7 when $\lambda = 2.61$, and to S_6 and S_7 when $\lambda = 7.83$. Since $\partial T_i/\partial\lambda_i$ is unbounded, γ_i is unbounded, and there is no overflow.

1.4 Performance Comparison

For $\lambda = x\lambda_{\max}$, where $x = 0.45, 0.55, 0.65, 0.75, 0.85$, we show in Tables 3 and 4 (for the idle-speed model and the constant-speed model respectively): (1) the optimal load distribution $\lambda_1, \lambda_2, ..., \lambda_n$; (2) ρ_i, $\bar{s}_i$, T_i, P_i, R_i, and $\rho(\lambda_1, \lambda_2, ..., \lambda_n)$, $\bar{s}(\lambda_1, \lambda_2, ..., \lambda_n)$, $T(\lambda_1, \lambda_2, ..., \lambda_n)$, $P(\lambda_1, \lambda_2, ..., \lambda_n)$, $R(\lambda_1, \lambda_2, ..., \lambda_n)$, with dynamic speed and power management. It is observed that the cost-performance ratio is indeed minimized. As an example, when $\lambda = 33.93$, for the idle-speed model, we have from Table 2 (of the main paper) that $T(\lambda_1, \lambda_2, ..., \lambda_n) = 1.4118793$ and $P(\lambda_1, \lambda_2, ..., \lambda_n) = 24.7737396$, which give $R(\lambda_1, \lambda_2, ..., \lambda_n) = 34.9775301$, and from Table 6 (of the main paper) that $T(\lambda_1, \lambda_2, ..., \lambda_n) = 1.5114816$ and $P(\lambda_1, \lambda_2, ..., \lambda_n) = 23.6305970$, which give $R(\lambda_1, \lambda_2, ..., \lambda_n) = 35.7172126$. In Table 3, R is reduced to 33.5965375 with $T = 1.4304299$ and $P = 23.9791773$. Also, for the constant-speed model, we have from Table 3 (of the main paper) that $T(\lambda_1, \lambda_2, ..., \lambda_n) = 1.4118793$ and $P(\lambda_1, \lambda_2, ..., \lambda_n) = 26.0853500$, which give $R(\lambda_1, \lambda_2, ..., \lambda_n) = 36.8293657$, and from Table 7 (of the main paper) that $T(\lambda_1, \lambda_2, ..., \lambda_n) = 1.4904961$ and $P(\lambda_1, \lambda_2, ..., \lambda_n) = 25.0270766$, which give $R(\lambda_1, \lambda_2, ..., \lambda_n) = 37.3027601$. In Table 4, R is reduced to 35.4442924 with $T = 1.4298174$ and $P = 25.3099391$. It is clear that the cost-performance ratio is minimized by finer balancing between cost (i.e., power consumption) and performance (i.e., the task response time), and in particular, by making T to be greater than the T obtained by minimizing the average task response time but less than the T obtained by minimizing the average power consumption, and by making P to be greater than the P obtained by minimizing the average power consumption but less than the P obtained by minimizing the average task response time.

TABLE 3
Numerical Data for Cost-Performance Ratio Minimization (Idle-Speed Model)

i	λ_i	ρ_i	$\bar{s}_i$	T_i	P_i	R_i
			$\lambda = 23.49$			
1	2.4361946	0.7461769	1.0658879	1.2930601	12.3548680	15.9755864
2	2.3094640	0.7366842	1.0331371	1.3986001	11.2717472	15.7646664
3	3.6685186	0.6903198	1.0433839	1.0594048	18.7947369	19.9112348
4	3.4568634	0.6777971	1.0135756	1.1320677	17.3362078	19.6257614
5	3.2927648	0.6551823	1.0033706	1.1533351	16.6774952	19.2347409
6	4.2104035	0.5992270	1.0022592	1.0464224	22.5153760	23.5605943
7	4.1157912	0.5876767	1.0002935	1.0504316	22.2424607	23.3641846
	Average	0.6595906	1.0197934	1.1369473	18.1467650	20.2176700
			$\lambda = 28.71$			
1	2.8755181	0.8338839	1.1246221	1.4324692	14.5550336	20.8496381
2	2.7437693	0.8360535	1.0785363	1.5936109	13.1034042	20.8817283
3	4.4701536	0.7987612	1.0952695	1.1242994	22.1413621	24.8935203
4	4.2275181	0.8004973	1.0450063	1.2451123	19.9102284	24.7904704
5	3.9348411	0.7723433	1.0146249	1.3077789	18.2843458	23.9118815
6	5.3616528	0.7505330	1.0154174	1.1321053	25.3686773	28.7200141
7	5.0965470	0.7248371	1.0032411	1.1516711	24.3142561	28.0020259
	Average	0.7803483	1.0469070	1.2492693	20.6490737	25.2218049
			$\lambda = 33.93$			
1	3.3395905	0.9038085	1.2093883	1.6752686	17.3904181	29.1336212
2	2.9999938	0.8818462	1.1181518	1.7223388	14.4971557	24.9690140
3	5.2396723	0.8806079	1.1673266	1.2644575	26.1015175	33.0042593
4	4.9999804	0.8920089	1.1079871	1.3856997	23.6711402	32.8010924
5	4.7337467	0.8896784	1.0570709	1.5881310	21.2125288	33.6882739
6	6.5224936	0.8736318	1.0581530	1.2877508	29.4792735	37.9619572
7	6.0945225	0.8508135	1.0198325	1.3621739	26.9477195	36.7074793
	Average	0.8792538	1.0955122	1.4304299	23.9791773	33.5965375
			$\lambda = 39.15$			
1	3.6568482	0.9397549	1.2791945	1.9906060	19.5955718	39.0070622
2	3.7425946	0.9639925	1.2835391	2.2273612	19.9523369	44.4410612
3	5.7341074	0.9226783	1.2241432	1.4642324	28.9994256	42.4618976
4	6.0440143	0.9630733	1.2457296	1.6312877	31.1027576	50.7375447
5	5.6137959	0.9639900	1.1587692	1.9305180	26.6486637	51.4457237
6	7.3329684	0.9340784	1.1134885	1.4824861	33.4165669	49.5395946
7	7.0256713	0.9365320	1.0671353	1.6649918	30.7251433	51.1571112
	Average	0.9450042	1.1800198	1.7184492	28.3808535	47.7802919
			$\lambda = 44.37$			
1	3.9766786	0.9678624	1.3576971	2.7020803	22.0015418	59.4499328
2	3.8999954	0.9736645	1.3263340	2.4043196	21.3562047	51.3471424
3	6.2185491	0.9569354	1.2867744	1.9319299	32.0727170	61.9622395
4	6.4999963	0.9793190	1.3206802	1.7994250	35.1554580	63.2596092
5	7.3050355	0.9981972	1.4628099	2.4649893	44.3586496	109.3435956
6	8.0760285	0.9711874	1.1825310	1.8962957	37.7928034	71.6663318
7	8.3937166	0.9907680	1.2083344	2.2851781	39.8101404	90.9732624
	Average	0.9784520	1.2967446	2.1711676	35.2073603	76.0493874

TABLE 4
Numerical Data for Cost-Performance Ratio Minimization (Constant-Speed Model)

i	λ_i	ρ_i	$\bar{s}_i$	T_i	P_i	R_i
			$\lambda = 23.49$			
1	2.4324395	0.7453337	1.0654794	1.2920625	13.8661637	17.9159506
2	2.3056368	0.7356961	1.0328495	1.3970155	12.8440165	17.9432901
3	3.6968282	0.6945561	1.0448095	1.0609899	21.9536938	23.2926467
4	3.4784685	0.6815764	1.0141173	1.1347789	20.5808975	23.3547681
5	3.2813371	0.6529954	1.0032721	1.1512112	20.1219507	23.1646143
6	4.2152945	0.5999035	1.0022815	1.0466704	28.1274419	29.4401621
7	4.0799954	0.5825914	1.0002651	1.0481797	28.0135329	29.3632174
	Average	0.6596919	1.0200600	1.1366355	21.9233172	24.3597092
			$\lambda = 28.71$			
1	2.8511991	0.8295931	1.1208065	1.4232202	15.4429867	21.9787701
2	2.7264940	0.8326145	1.0762168	1.5853115	14.0226346	22.2302439
3	4.4533683	0.7967373	1.0939364	1.1223715	24.0956266	27.0442458
4	4.2315910	0.8010693	1.0452489	1.2457897	21.9158718	27.3025677
5	3.9298124	0.7714844	1.0144781	1.3062970	20.5551251	26.8510991
6	5.4068518	0.7560100	1.0163974	1.1366856	28.9159594	32.8683553
7	5.1106835	0.7267588	1.0033389	1.1538069	28.1715871	32.5045710
	Average	0.7804623	1.0461398	1.2478707	23.1071933	28.1644877
			$\lambda = 33.93$			
1	3.3197387	0.9012593	1.2053203	1.6611657	17.8516282	29.6545133
2	2.9999938	0.8818462	1.1181518	1.7223388	15.2060788	26.1900198
3	5.2181232	0.8785965	1.1650281	1.2585359	27.1952263	34.2261687
4	4.9999804	0.8920089	1.1079871	1.3856997	24.7510508	34.2975242
5	4.7344858	0.8897662	1.0571310	1.5884203	22.3183278	35.4508853
6	6.5377437	0.8749728	1.0589906	1.2904762	31.2952126	40.3857255
7	6.1199344	0.8536554	1.0206210	1.3691714	29.0781969	39.8130349
	Average	0.8794394	1.0948861	1.4298174	25.3099391	35.4442924
			$\lambda = 39.15$			
1	3.6501264	0.9390818	1.2776270	1.9814631	19.9123706	39.4556271
2	3.7341058	0.9634087	1.2812933	2.2190801	20.0983822	44.5999195
3	5.7275855	0.9221722	1.2233449	1.4604783	29.7378387	43.4314668
4	6.0334231	0.9626082	1.2440764	1.6281121	31.3879199	51.1030507
5	5.6115765	0.9638729	1.1584424	1.9297090	26.9927209	52.0880969
6	7.3400449	0.9345104	1.1140675	1.4848511	34.3718092	51.0370181
7	7.0531378	0.9384259	1.0691652	1.6751679	31.7269928	53.1480395
	Average	0.9451321	1.1795607	1.7177458	28.9903439	48.7717344
			$\lambda = 44.37$			
1	3.9763021	0.9678336	1.3576005	2.7007246	22.1916134	59.9334358
2	3.8999954	0.9736645	1.3263340	2.4043196	21.5142179	51.7270566
3	6.2188579	0.9569553	1.2868163	1.9324218	32.5051908	62.8137381
4	6.4999963	0.9793190	1.3206802	1.7994250	35.3622675	63.6317475
5	7.2993685	0.9981760	1.4616977	2.4628990	44.3048018	109.1182540
6	8.0790354	0.9713061	1.1828418	1.8991587	38.2135441	72.5735837
7	8.3964443	0.9908125	1.2086795	2.2869170	39.9607572	91.3869344
	Average	0.9784764	1.2966458	2.1715662	35.4248277	76.5027896