

Системы с инверсионным обслуживанием и обобщенным вероятностным приоритетом и их применение к оценке показателей эффективности систем распределенных вычислений

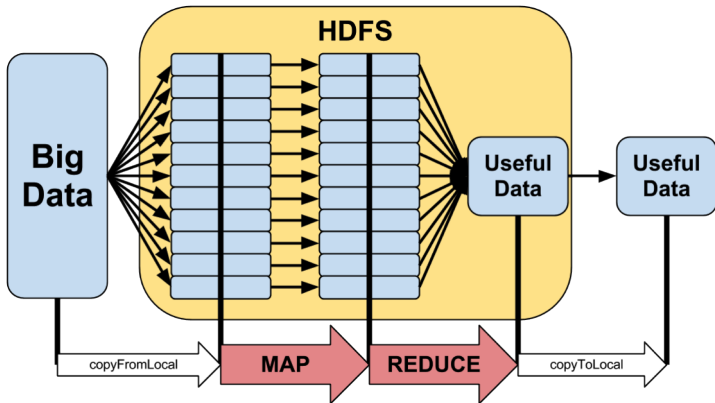
по материалам диссертации на соискание ученой степени кандидата
физико-математических наук по специальности
05.13.17 – теоретические основы информатики

Л. А. Мейханаджян

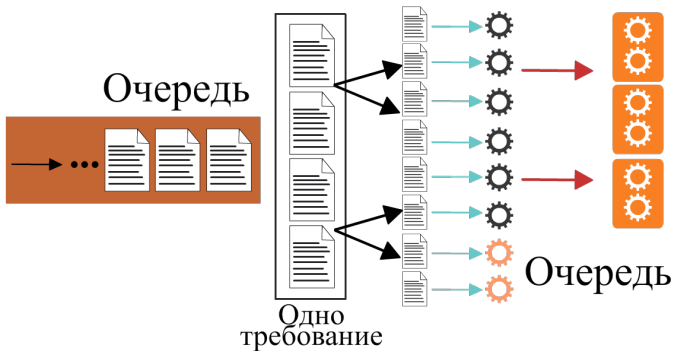
Научный руководитель: к.ф.-м.н., Р. В. Разумчик

Москва, 2016

Hadoop I



Hadoop II



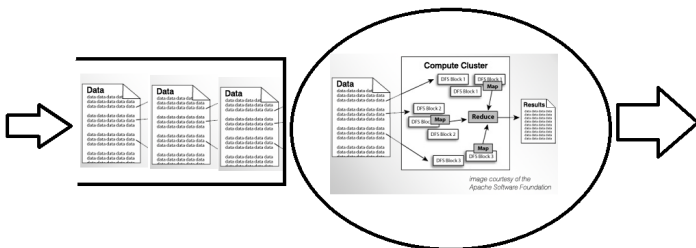
Hadoop \rightarrow M|G|1 I

Основные предположения:

- ✓ каждое задание определяется всего двумя числами: время поступления в систему, время обслуживания;
- ✓ число задач в каждом задании таково, что задание может занять все рабочие машины кластера одновременно;
- ✓ прерывание задания и расположение данных не влияет на общее время выполнения задания;
- ✓ время выполнения задания известно неточно.

Hadoop → M|G|1 II

- Показатели эффективности
 - ✓ быстрое время отклика
 - ✓ slowdown
- Используемые дисциплины обслуживания
 - ✓ HFS
 - ✓ FIFO
 - ✓ HSBP



Постановка задачи I

- Существуют современные системы обработки данных, в которых времена выполнения заданий известны неточно (например, система распределенной обработки данных Hadoop) пример [1,2]
- Для повышения производительности таких систем используются специальные дисциплины обслуживания (например, SRPT + PS) [3]
Эти специальные дисциплины отличаются от классических!
- Для обоснования эффективности новых специальных дисциплин (типа HFSP), используются методы ТМО (например, Hadoop \rightarrow M/G/1 \rightarrow HFSP) [4,5]

Постановка задачи II

Пусть имеется адекватная математическая модель системы с неточной информацией о временах выполнения заданий (например, Hadoop \rightarrow M/G/1).

Как значения характеристик производительности матмодели отличаются от значений характеристик реальной системы?

Можно ли внося изменения лишь в математическую модель, получить более близкие к практике результаты?

Можно, если распределение времени обслуживания имеет тяжелый хвост!

Время пребывания I

Предполагается, что имеет место соотношение

$$Z = X \cdot Y$$

где X – фактическое время выполнения заданий, Z – прогнозное время выполнения заданий.

Обозначим через $V(x)$ ФР прогнозного времени выполнения задания, а через $G(x)$ – ФР фактического времени выполнения задания.

Время пребывания II

Используя известные результаты для систем $M|G|1$, с дисциплинами FCFS, PS, PLCFS либо SPLCFS, можно рассчитать среднее время отклика v_B , соответствующее прогнозируемому времени выполнения задания при дисциплине FCFS

$$v_B = EX + \frac{\lambda EX^2}{2(1 - \lambda EX)},$$

где λ – интенсивность входящего потока, случайная величина X есть длительность обслуживания заявки на приборе, EX^2 – 2-й момент времени обслуживания.

При дисциплине PS и pLCFS равно

$$\frac{EX}{1 - \lambda EX},$$

а при дисциплине spLCFS равно

$$\frac{EX(2 - EX)}{2(1 - \lambda EX)}.$$

Формулировка

Имеется модель реальной системы в виде СМО M/G/1 с дисциплиной обслуживания PS/FIFO/LIFO/Pre Size-Based

- ✓ интенсивность входящего потока известна;
- ✓ ФР времени обслуживания известна, имеет тяжелый хвост, содержит ошибку (то есть отличается от ФР, которая будет наблюдаться в реальной системе).

Можно ли, изменяя лишь правила обслуживания заявок, получить оценку v^* для среднего времени отклика, удовлетворяющую неравенствам

$$v_G < v^* < v_B.$$

Описание системы I

Обобщенный инверсионный порядок обслуживания с вероятностным приоритетом (LCFS BPP) заключается в следующем

- с вероятностью $D(x, y|u, v)$ обслуживавшаяся ранее заявка продолжает обслуживаться, причем ее длина становится меньше y , а вновь поступившая становится на первое место в очереди и ее длина становится меньше x , а с вероятностью $D^*(x, y|u, v)$ наоборот;
- с вероятностью $D_0(x|u, v)$ заявка, находящаяся на приборе, покидает систему, а поступившая заявка становится на прибор, причем ее длина становится меньше x , а с вероятностью $D_0^*(y|u, v)$ наоборот;
- с вероятностью $d_0(u, v)$ обе заявки покидают систему, а на прибор становится первая заявка из очереди.

Стационарные вероятности I

$$-p'_n(x) = a_n(x) - \lambda p_n(x) + \int_0^{\infty} K_n(x, v) p_n(v) dv, \quad n \geq 1,$$

с начальным условием

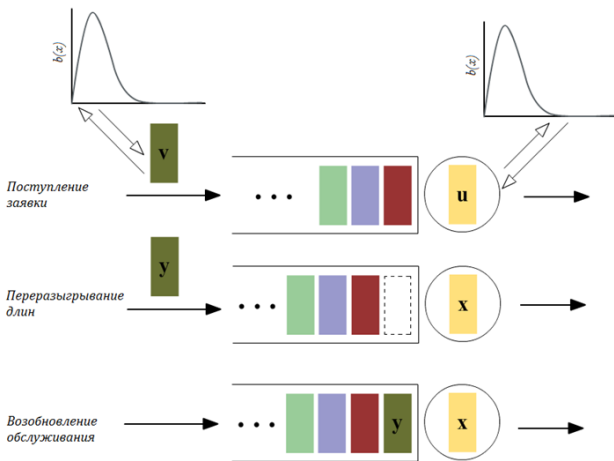
$$p_n(\infty) = \lim_{X \rightarrow \infty} p_n(X) = 0, \quad n \geq 1, \quad \text{где}$$

$$a_1(x) = \tilde{\lambda} \tilde{b}(x) p_0,$$

$$a_n(x) = \lambda \left(\int_0^{\infty} p_{n-1}(v) dv \int_0^{\infty} b(u) du \int_0^{\infty} [d(y, x|u, v) + d^*(x, y|u, v)] dy \right), \quad n \geq 2,$$

$$K_n(x, v) = \lambda \int_0^{\infty} b(u) du \int_0^{\infty} d(x, y|u, v) dy, \quad n \geq 1.$$

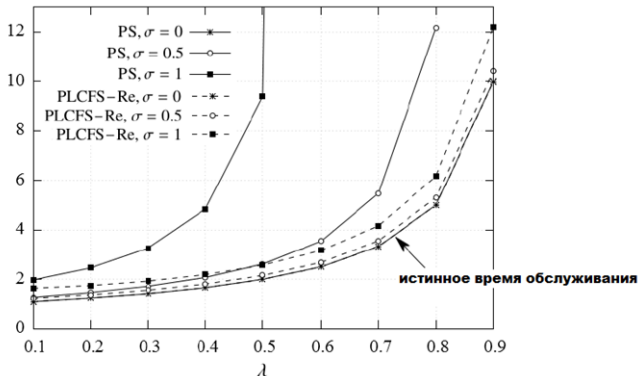
$$d(x, y|u, v) = b(x)b(y)$$



Примеры I

Исходные данные:

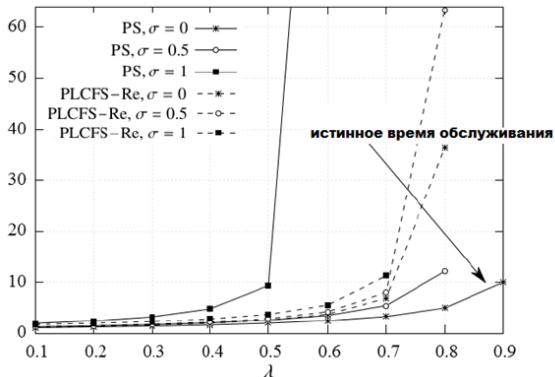
- ✓ Дисциплина обслуживания PS;
- ✓ Истинное время обслуживания S имеет экспоненциальное распределение;
- ✓ Доступна информация только о $X \cdot S$, где X имеет логнормальное распределение с параметрами $(0, \sigma^2)$;
- ✓ Интересует оценка среднего времени пребывания задания в системе.



Примеры II

Исходные данные:

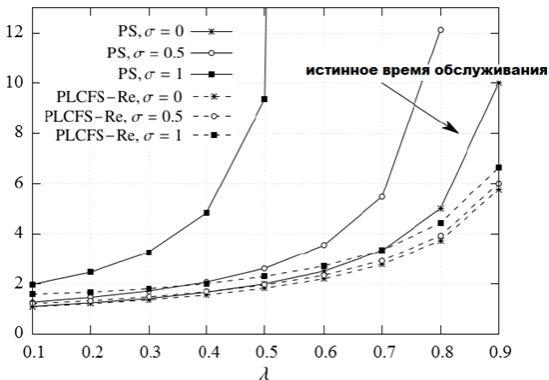
- ✓ Дисциплина обслуживания PS;
- ✓ Истинное время обслуживания S имеет распределение с легким хвостом;
- ✓ Доступна информация только о $X \cdot S$, где X имеет логнормальное распределение с параметрами $(0, \sigma^2)$;
- ✓ Интересует оценка среднего времени пребывания задания в системе.



Примеры III

Исходные данные:

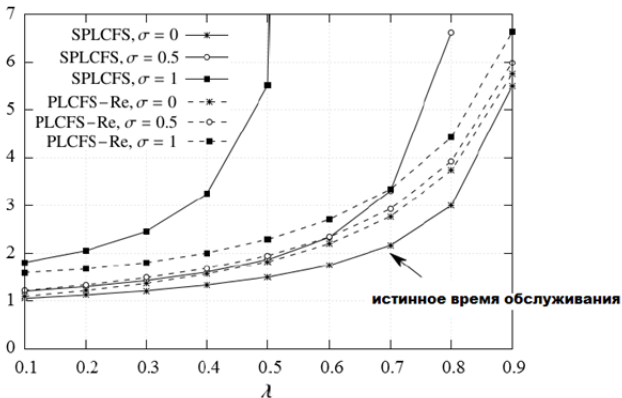
- ✓ Дисциплина обслуживания PS;
- ✓ Истинное время обслуживания S имеет распределение с длинным хвостом (30/70);
- ✓ Доступна информация только о $X \cdot S$, где X имеет логнормальное распределение с параметрами $(0, \sigma^2)$;
- ✓ Интересует оценка среднего времени пребывания задания в системе.



Примеры IV

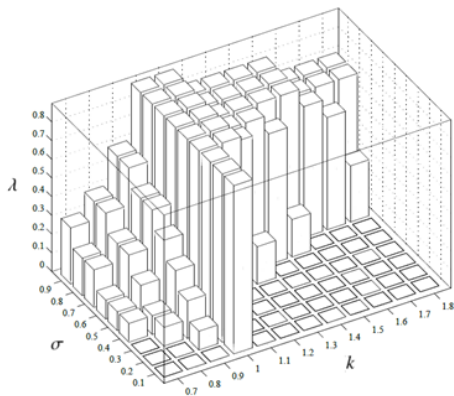
Исходные данные:

- ✓ Дисциплина обслуживания S-PLCFS;
- ✓ Истинное время обслуживания S имеет распределение с длинным хвостом (30/70);
- ✓ Доступна информация только о $X \cdot S$, где X имеет логнормальное распределение с параметрами $(0, \sigma^2)$;
- ✓ Интересует оценка среднего времени пребывания задания в системе.



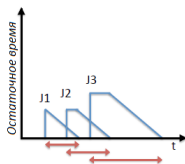
Всегда ли можно получить оценку, действуя таким образом?

Нет! Все зависит от свойств распределения времени обслуживания!

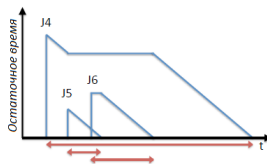
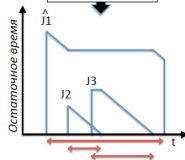


СПИСОК ИСТОЧНИКОВ ●

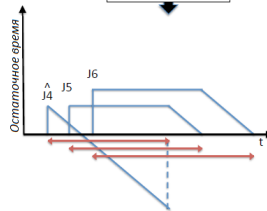
1. Verma A., Cherkasova L., Campbell R. H. Two sides of a coin: Optimizing the schedule of MapReduce jobs to minimize their makespan and improve cluster performance // Proc. IEEE 20th Int. Symp. Model., Anal. Simul. Comput. Telecommun. Syst., 2012. P.11–18.
2. Dell’Amico M. A simulator for data-intensive job scheduling // arXiv, Tech. Rep. arXiv:1306.6023. 2013.
3. Dell’Amico M., Carra D., Michiardi P. PSBS: Practical Size-Based Scheduling // IEEE Trans. Computers, 2016. Vol. 65. Iss. 7. P. 2199–2012.
4. Pastorelli M., Barbuzzi A., Carra D., Dell’Amico M., Michiardi P. HFSP: Bringing Size-Based Scheduling for Hadoop // IEEE Big Data, 2013. P. 51–59.
5. Wierman A., Nuyens M. Scheduling despite inexact job-size information // ACM SIGMETRICS Performance Evaluation Review. 2008. Vol. 36. Iss. 1. P. 25–36.




Переоценена



Недооценена



Пример планирования при наличии неточной информации о времени обслуживания: с ошибками (вверх) и без (низ). Пример взят из Dell'Amico M, Carra D, Pastorelli M, Michiardi P. 2014. Revisiting size-based scheduling with estimated job sizes. Available in arxiv.org, arXiv:1403.5996. 

σ	$k = 0.7$		$k = 1$		$k = 1.8$	
	joint ratio $p\%/(1-p)\%$	m-m dist.	joint ratio $p\%/(1-p)\%$	m-m dist.	joint ratio $p\%/(1-p)\%$	m-m dist.
0	26.55/73.45 ^a	4.87	31.92/68.08 ^b	2.42	38.68/61.32 ^c	1.38
0.1	26.49/73.51	4.91	31.86/68.14	2.44	38.48/61.52	1.39
0.2	26.31/73.69	5.05	31.49/68.51	2.52	38.01/61.99	1.43
0.3	25.89/74.11	5.33	30.98/69.02	2.65	37.39/62.61	1.51
0.4	25.44/74.56	5.68	30.45/69.55	2.84	36.43/63.57	1.63
0.5	24.93/75.07	6.25	29.55/70.45	3.12	35.13/64.87	1.78
0.6	24.24/75.76	7.00	28.69/71.31	3.50	33.91/66.09	1.99
0.7	23.52/76.48	8.01	27.69/72.31	4.00	32.50/67.50	2.28
0.8	22.72/77.28	9.34	26.71/73.29	4.66	31.04/68.96	2.66
0.9	21.90/78.10	11.13	25.60/74.40	5.55	29.59/70.41	3.16