

МЕТОДИЧЕСКИ ПОДХОД ЗА ОПТИМИЗИРАНЕ НА РАЗХОДИТЕ В CLOUD COMPUTING

Гроздалина Николаева Грозева

Department of Management, Technical University, Kliment Ohridski 8, 1000 Sofia,
Bulgaria, E-mail: grozeva77@abv.bg

Keywords: algorithm, data centre, cloud computing

Abstract Following the increasing number of providers of Cloud services and the variety of different pricing schemes, the question arises how to create competitive advantage by minimizing costs and obtaining a larger income for suppliers. To achieve this goal several data centers are considered that are connected to each other by common infrastructure based on a cluster and the physical resources located at the centers. In this report a new methodological approach for dynamic resource management in Cloud environments is presented, in order to obtain optimal cost.

Резюме С нарастване на броя на доставчиците на Cloud услуги и многообразието от различни ценови схеми възниква въпроса за създаване на конкурентно предимство, чрез минимизиране на разходите и получаване на по - голям доход за доставчиците. За постигането на тази цел са разгледани няколко центъра за данни, които са свързани с обща инфраструктура по между си на базата на клъстер и със съответните физически ресурси разположени в тях. В настоящият доклад е представен нов методически подход за динамично управление на ресурсите в Cloud среда, с който да се получи оптимизиране на разходите.

1. ВЪВЕДЕНИЕ

Центровете за данни (Cloud), традиционно са домакини на голям брой различни услуги, като IaaS, PaaS, SaaS [1, 2, 3] с обща инфраструктура. Доставчиците на инфраструктура предлагат персонализиран капацитет с различни хардуерни конфигурации, операционни системи и допълнителни услуги. Отделните доставчици предлагат различни ценови схеми [2, 3] за използване на специфичните си инфраструктурни ресурси. Управлението на ресурсите по заявка в центровете за данни е необходимо, за да максимизира приходите на доставчиците, като отговаря на изисканото качество за обслужване и е насочено към минимизиране на оперативните разходи. С нарастване на броя на доставчиците на Cloud услуги и многообразието от различни ценови схеми възниква въпросът за създаване

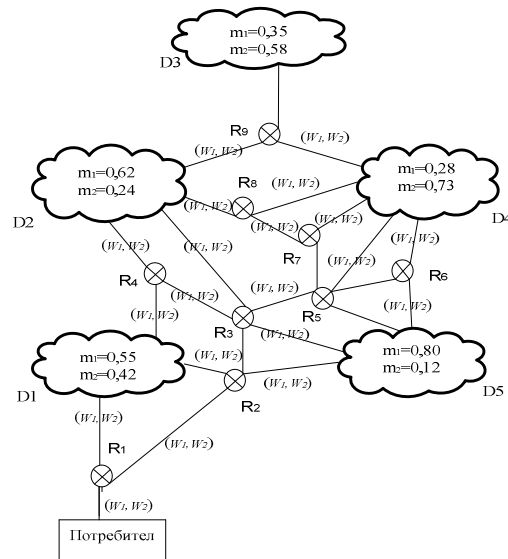
на конкурентно предимство [4, 5, 6], чрез търсене и намиране на решение за икономическа ефективност т.е. минимизиране на разходите и получаване на по - голям доход за доставчиците.

Този доклад има за цел да представи нов методически подход за динамично управление на ресурсите в Cloud среда, с който да се получи желания резултат. Доставчикът на инфраструктура предлага ресурси под наем съгласно специфична ценова схема, заедно с управленски интерфейс за потребителите, за преглед (browse), придобиване, мониторинг и контрол на ресурсите. Ресурсите могат да бъдат: изчислителен ресурс с избрана операционна система, ресурс за съхранение или предоставяне на услуга с предефинирана функционалност. За доставчиците, изчислителният ресурс е виртуална машина (VM) със специфична хардуерна конфигурация и тип операционна система, но може да съдържа и друг софтуер или данни, заявени от потребителя. Доставчикът на инфраструктура предлага различни хардуерни конфигурации с вариращи ценови модели. Всеки физически инфраструктурен ресурс е собственост на доставчика на инфраструктура и следва да бъде споделен между потребителите.[7]

Изчислителните платформи традиционно имат включени кълстери и изчислителни мрежи [6, 8], каквато е Cloud, появила се като иновация в телекомуникациите, която е разходно ефективна и мощна платформа. Иновациите в телекомуникационния сектор са необходимост, а не опция, тъй като цялата макро икономика зависи от създаването на динамична, конкурентноспособна, иновативна телекомуникационна инфраструктура. За много компании качеството и цената на телекомуникационните услуги са едни от най-важните съображения за избор на локация.

На фигура 1 е представена блокова схема на доставчик на Cloud услуги. Тези услуги са разпределени в няколко центъра за данни на базата на кълстер. Схемата се състои от две основни части. Първата част представлява различните центрове за данни, отбелязани с D_i , където $i=1 \div 5$ със съответните сървъри, разположени в тях. Втората част е мрежовата инфраструктура, където с R_i са отбелязани съответните маршрутизатори и линкове (връзки), които са физическата връзка между тях. Посочен е също така един потребител, който е случайно разположен. В схемата са зададени две случайни стойности на всеки един център за данни, цена за единица ресурс и натовареност, изразена в проценти, което представлява моментната използваемост на капацитета на съответния

център за данни със съответния разход за ресурси. Всеки един център за данни е позициониран в различна държава и съответно има различна себестойност на единица използван ресурс.



Фиг.1. Примерна блокова схема на доставчик на Cloud услуги.

2. МЕТОДИЧЕСКИ ПОДХОД

Целта, която си поставяме е свързана с избор на център за данни с минимална натовареност и минимална цена за единица ресурс. На фигура 1 се вижда, че избраният потребител се намира в най - голяма близост до център за данни D_1 . Това, обаче, не е оптималното решение. Когато параметърът е един, решението е тривиално. В нашия случай, параметрите са два, което значително усложнява задачата. Поради тази причина се предлага подход, при който се миксират двата параметъра чрез теглови коефициенти, което води до получаването на един параметър.

$$m = k_1 m_1 + k_2 m_2, \quad (1)$$

където:

m - миксиран параметър, m_1 - натовареност (изчислителна мощност, операциона памет), m_2 - цена (лв/единица ресурс), k_1 и k_2 - съответните теглови коефициенти. Тегловите коефициенти се избират динамично в зависимост от моментното състояние на Cloud средата. За по-голяма яснота можем да разгледаме следните два примера.

Пример 1: Избираме $k_1 = 2, k_2 = 1$

$$D_1 = 1.52, D_2 = 1.48, D_3 = 1.28, D_4 = 1.29, D_5 = 1.72$$

От тук следва че D_3 е оптималният избор

Пример 2: $k_1 = 1, k_2 = 2$

$$D_1 = 1.39, D_2 = 1.10, D_3 = 1.51, D_4 = 1.74, D_5 = 1.04$$

При така поставените условия D_5 е оптималният избор. Въпросът, който възниква веднага е възможно ли е всеки един потребител да използва център за данни, избран по посочения метод. В нашия случай, потребителят се намира най - близко до център за данни D_1 , а оптималните решения в посочените примери по-горе са D_3 и D_5 . За да бъде възможно това, трябва да бъдат изпълнени две условия:

1. Техническите. параметри на мрежата да отговарят на изискваното Качество за Обслужване на даденото приложение.
2. Цената на D_i плюс цената на мрежата да е по – малка от D_1 .
Следователно за цената на мрежата C_n трябва да е изпълнено (2)

$$C_n < C_{D_1} - C_{D_i}, \quad (2)$$

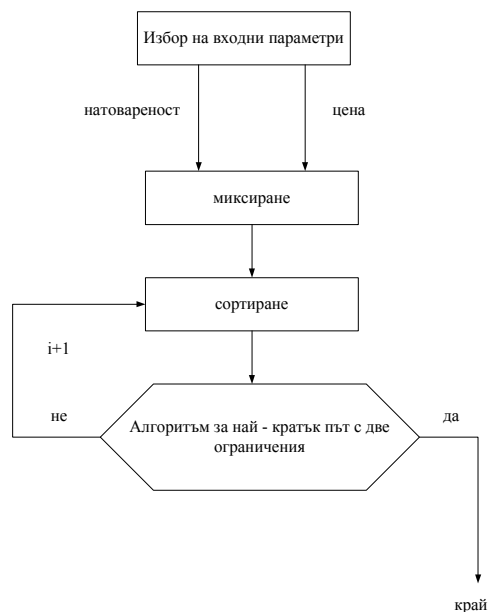
където:

C_{D_1} - цената в най-близкия дейта център, в нашия случай това е D_1

C_{D_i} - цената в избрания оптимален център за данни.

Тези две условия поставят нова задача, която е свързана с намиране на най-кратък път с две ограничения от дадения потребител до оптималния център за данни. Тази задача ще бъде подробно разгледана в точка 3.

На фигура 2 е дадена блок схема на предложеният от нас алгоритмичен подход.



Фигура 2. Алгоритмичен подход

На първия етап на алгоритъма за определяне на входните данни бе избран разширения модел за TCO (Total Cost of Ownership) на [2, 3], приложен за IaaS, PaaS, и SaaS. TCO е фокусиран върху отделни характеристики на тези модели на услуги и включва широк кръг на разходни типове и фактори. Той може да се използва като инструмент за оптимизация на разходите. Методиката TCO позволява да се получи адекватна информация, да се направят сравнения и да се оптимизират инвестициите на компанията. TCO включва преки и косвени разходи. Въз основа на този модел са избрани технически характеристики като: изчислителна мощност (CPU), операционна памет, съхранение. Тези характеристики определят техническият параметър натовареност, а за икономически параметър избираме – цена (лева за единица ресурс) .

На втория етап от алгоритъма следва миксиране на двата параметъра – цена и натовареност, които са изчислени по посочения метод. След като сме определили миксирания параметър за всеки център за данни, се сортира във възходящ ред. Когато сортирането приключи, прилагаме алгоритъм за намиране на най - кратък път с две ограничения, който ще разгледаме по - подробно в т.3.

Ако са изпълнени двете ограничителни условия - следва край, ако не - алгоритъмът проверява следващия по ред център за данни.

3. АЛГОРИТЪМ ЗА НАЙ - КРАТЪК ПЪТ С ДВЕ ОГРАНИЧЕНИЯ

Всеки един линк (фигура 1) е натоварен с двуменсионен линк вектор $\vec{w}(w_1, w_2)$, където w_1 е цената, докато w_2 е технически параметър, изискван от дадено приложение, за да бъде гарантирано дадено качество на обслужване.

В литературата съществуват множество алгоритми, които са способни да решат задачата за най-кратък път, когато параметърът е само един. Най-популярни са алгоритмите на Белман – Форд и Дикстра [9,10]. Те обаче не могат да решат задачата, когато параметрите са два или повече от два. Всеки един параметър е сума от теглата му върху всеки линк по целия път от източника до получателя (3)

$$w_i(e) = \sum_{e \in p} w_i(e), \quad (3)$$

където, p - пътят от източника до получателя, а e - съответните линкове.

За да бъде възможно използването на получения оптимален център за данни, трябва да са изпълнени условията (4) и (5)

$$C_n = w_1(p) < C_{D_1} - C_{D_i}, \quad (4)$$

$$w_2(p) < L, \quad (5)$$

където:

$w_1(p)$ - цената на целия път,

$w_2(p)$ - изискваният технически параметър,

L - границата на изисквания от приложението технически параметър.

Така поставена задачата е известна като Multi Constraint Problem (MCP). MCP е NP hard (нерешим в полиномно време) [11]. В литературата съществуват алгоритми, които решават този вид задачи [12,13,14,15,16,17,18,19]. С имплементирането на един от тези алгоритми ние постигаме решаването на поставената от нас задача.

4. ЗАКЛЮЧЕНИЕ

В тази статия е разработен нов методически подход за оптимизиране на разходите в Cloud среда. Целта, която бе поставена е свързана с избор на център за данни с минимална натовареност и минимална цена за единица ресурс. Когато параметърът е един, решението е тривиално. В нашия случай, параметрите са два, което значително усложнява задачата. Поради тази причина се предлага подход, при който се миксират двата параметъра чрез теглови коефициенти. След като сме определили миксираната метрика за всеки център за данни, се сортират във възходящ ред. Когато сортирането приключи, прилагаме алгоритъм за намиране на най - кратък път с две ограничения. За постигане на целта трябваше да бъдат изпълнени двете условия. Така поставената задача е известна като MCP (Multi Constraint Problem), която е нерешима в полиномно време. С имплементирането на един от предложените алгоритми за намиране на най-кратък път с две ограничения се постигна решаването на поставената задача.

За бъдеща работа предстои да се направят експериментални изследвания. Получените резултати да се анализират и да се направи сравнение с резултатите към момента в Cloud среда, работеща на случаен принцип, както и да се оценят предимствата и недостатъците.

БЛАГОДАРНОСТИ

Докладът е разработен по проект 122ПД0064-15/24.04.2012г.”Изследване и разработване на методически подход за икономическа оценка на иновации в телекомуникациите.”

REFERENCE

1. Armbrust, M. Fox, A. Griffith, R. Joseph, A D. Katz, , Konwinski, A. Lee, G. Patterso, D. Rabkin, A. Stoica, I. and Zaharia, M. Above the Clouds: A Berkeley View of Cloud Computing 2009
2. B. Martens, M. Walterbusch, and F. Teuteberg, „ Costing of Cloud Computing Services: A Total Cost of Ownership Approach” 45th Hawaii International Conference on System Sciences 2012, pp.1563-1572, IEEE Computer Society

3. J. Strebel, A. Stage „ An economic decision model for business software application deployment on hybrid Cloud environments. Multiconferenz Wirtschaftsinformatik. 2010
4. .Портър, М., Конкурентно предимство на нациите С. 2004
5. Велев. М, Оценка и анализ на фирмената конкурентоспособност, С., 2004..
6. Фондация „Приложни изследвания и комуникации” Клъстери и иновации – възможна парадигма на обновлението 2008 pp 309-337
7. P. Wright, Y.Sun, T. Harmer, A. Keenan, A. Stewart, R. Perrott, „ A constraints-based resource discovery model for multi-provider Cloud environment” Journal of Cloud Computing: Advances, Systems and Applications 21. June 2012
8. D.Kondo, B.Javadi, P. Malecot, F.Cappello, D. Anderson „ Cost – Benefit Analysis of Cloud Computing versus Desktop Grids” INRIA, France, UC Berkeley, USA
9. R. Bellman , On a Routing Problem. Quarterly Applied Mathematics, 1958 XVI(1), 87-90
10. E.W. Dijkstra. A note on two problems in connexion with graphs. Numerische Mathematik, (1):269—271, 1959.
11. M.R. Garey and D.S. Johnson. Computers and Intractability: A guide to the Theory of NP-Completeness. ISBN 0-7167-1044-7. W.H. Freeman and Company, San Francisco, 1979
12. J.M. Jaffe. Algorithms for finding paths with multiple constraints. Networks, 14:95—116, 1984
13. A. Iwata, R. Izmailov, D. Lee, B. Sengupta, G. Ramamurthy, and H. Suzuki. ATM routing algorithms with multiple QoS requirements for multimedia internetworking. IEICE Trans. Commun., E79-B(8):999—1007, Aug. 1996
14. A. Juttner, B. Szviatovszki, I. Mecs, Z. Rajko, “Lagrange relaxation based method for the QoS routing problem”, Proceedings of the INFOCOM 2001 Conference, vol. 2, IEEE, 2001, pp. 859–868
15. Gang Feng, “ Exact algorithms for multi-constrained QoS routing ,” International Conference on Computer, Communication and Control Technologies (CCCT’03), Orlando, USA, Vol. 2, pp. 340 – 345, July 31 – Aug 2, 2003
16. H. De Neve and P. Van Mieghem. TAMCRA: a tunable accuracy multiple constraints routing algorithm. Computer Communications, 23:667—679, 2000.
17. Van Mieghem P. and F. A. Kuipers, “Concepts of Exact Quality of Service Algorithms,” IEEE/ACM Trans. Net Vol. 12, pp. 851 - 862, 2004
18. T. Korkmaz, M. Krunz, “Multi-constrained optimal path selection”, Proceedings of the INFOCOM 2001 Conference, vol. 2, IEEE, Anchorage, Alaska, 2001, pp. 834–843.
19. Y. Tomov, G Iliev “A New Modified Algorithm for Multi Constraint Routing” ICEST 2012 June 28-30