Optymalizacja alokacji maszyn wirtualnych w klastrze obliczeniowym

Michał Karpowicz

Instytut Automatyki i Informatyki Stosowanej
ul. Nowowiejska 15/19, 00-665 Warszawa, Polska
m.karpowicz@elka.pw.edu.pl

Warszawa, Wrzesień 2018

Praca wykonana w ramach grantu OPUS Narodowego Centrum Nauki numer 2015/17/B/ST6/01885

Streszczenie
Raport prezentuje sformułowanie zadania optymalizacji, którego rozwiązaniem jest optymalna (w sensie zadanego kryterium) alokacja maszyn wirtualnych wykonywanych w klastrze obliczeniowym zbudowanym z wielu maszyn fizycznych.

Słowa kluczowe: Klastry obliczeniowe, alokacja maszyn wirtualnych, równoważenie obciążeń, optymalizacja.

1 Wprowadzenie

Model matematyczny składa się z dwóch zadań optymalizacji. Rozwiązaniem pierwszego z nich jest przydział maszyn wirtualnych do maszyn fizycznych klastra. Przydział ten jest optymalny z punktu widzenia zadanych kryteriów oceny rozwiązania oraz dopuszczalny względem wprowadzonych ograniczeń. Rozwiązaniem drugiego zadania optymalizacji jest alokacja zasobów obliczeniowych i zasobów pamięci maszyn fizycznych do umieszczonych na tych maszynach, na podstawie rozwiązania pierwszego zadania, maszyn wirtualnych. Rozwiązanie pierwszego zadania optymalizacji jest parametrem drugiego zadania optymalizacji. Przydział zasobów obliczeniowych i zasobów pamięci będący rozwiązaniem tego zadania maksymalizuje użyteczność zasobów fizycznych systemu z punktu widzenia świadczonych przez system usług.

W modelu uwzględniono zidentyfikowane eksperymentalnie profile zużycia zasobów systemowych przez maszyny wirtualne, a także wymagania dotyczące separacji wybranych usług systemu wynikające z przeprowadzonych badań diagnostycznych i realizowanych prac projektowych.

2 Formalny model wykorzystania zasobów systemu

Model matematyczny systemu ma postać grafu skierowanego złożonego z $m+n$ węzłów połączonych mn krawędziami. Węzły grafu reprezentują m maszyn wirtualnych realizujących ściśle określone usługi na n maszynach fizycznych. Krawędź łącząca węzeł $i = 1,\dots,m$ z węzłem $j = 1,\dots,n$ oznacza, że dopuszczalna jest alokacja maszyny wirtualnej i na maszynie fizycznej j. Zakłada się przy tym, że dowolna maszyna wirtualna może zostać umieszczona na dowolnej maszynie fizycznej.

Graf reprezentowany jest przez macierz incydencji (krawędź-węzeł) $A = [a_{ij}]_{(mn)\times(m+n)}$, gdzie $a_{ij} \in \{-1, 0, 1\}$. Węzły początkowy k oraz końcowy l krawędzi i są określone odpowiednio przez wyrazy $a_{ik} = -1$ oraz $a_{il} = 1$ wiersza i macierzy incydencji grafu.

2.1 Problem alokacji maszyn wirtualnych

Problem alokacji maszyn wirtualnych na maszynach fizycznych można sprowadzić do następującego zadania optymalizacji przepływów w grafie reprezentującym system:

$$\mathscr{P}[\mathbf{p},\mathbf{c},\mathbf{A},\underline{\mathbf{f}},\bar{\mathbf{f}},\mathbf{d},\underline{\mathbf{w}},\overline{\mathbf{w}}]:$$
$$\left|\begin{array}{ll} \text{zminimalizuj} & (\mathbf{p}^T\mathbf{A}^T + \mathbf{c}^T)\mathbf{w} \\ \text{przy ograniczeniach} & \begin{pmatrix} -\mathbf{A}^T \\ \mathbf{A}^T \\ \mathbf{S} \\ -\mathbf{I} \\ \mathbf{I} \end{pmatrix} \mathbf{w} \le \begin{pmatrix} -\underline{\mathbf{f}} \\ \bar{\bar{\mathbf{f}}} \\ \mathbf{d} \\ -\underline{\mathbf{w}} \\ \overline{\overline{\mathbf{w}}} \end{pmatrix}. \end{array}\right.$$

Rozwiązaniem powyższego zadania jest wektor $\hat{\mathbf{w}} = [w_i]_{mn\times 1} \in \{0,1\}^{mn}$ określający optymalny przydział m maszyn wirtualnych do n maszyn fizycznych klastra. Liczba maszyn wirtualnych przydzielonych do maszyn fizycznych określona jest przez wektor $\hat{\mathbf{u}} = \mathbf{A}^T\hat{\mathbf{w}}$ dywergencji węzłów grafu.

Wektor $\hat{\mathbf{w}}$ minimalizuje sumę kosztów związanych rozkładem ilości maszyn wirtualnych na węzłach fizycznych, $\mathbf{p}^T\hat{\mathbf{u}}$, oraz kosztu $\mathbf{c}^T\hat{\mathbf{w}}$ przydziału określonej maszyny wirtualnej do określonej maszyny fizycznej. Elementy wektora $\mathbf{p} = [p_j]_{(m+n)\times 1} \geq 0$ określają koszt cząstkowy przydziału pojedynczej maszyny wirtualnej do węzła fizycznego $j = 1, \ldots, n$, natomiast elementy wektora $\mathbf{c} = [c_i]_{mn\times 1} \geq 0$ koszt przydziału określonej maszyny wirtualnej do określonej maszyny fizycznej. Wektory $\mathbf{p}$ i $\mathbf{c}$ są parametrami zadania wyrażającymi preferencje administratora systemu dotyczące rozkładu maszyn wirtualnych w klastrze.

Wektor $\hat{\mathbf{w}}$ jest jednocześnie rozwiązaniem układu nierówności liniowych określających zbiór rozwiązań dopuszczalnych zadania. Nierówność:

$$\underline{\mathbf{f}} \leq \mathbf{A}^T\mathbf{w} \leq \bar{\mathbf{f}} \tag{1}$$

wskazuje najmniejszą i największą liczbę maszyn wirtualnych jaką można umieścić na maszynach fizycznych. Przyjmuje się przy tym, że $\underline{f}_j = \overline{f}_j = -1$ dla $j = 1, \ldots, m$. W taki sposób wymusza się alokację każdej maszyny wirtualnej na jednej z maszyn fizycznych.

Kolejna nierówność określa wymaganą przez administratora lub architekturę systemu separację maszyn wirtualnych:

$$\mathbf{S}\mathbf{w} \leq \mathbf{d}. \tag{2}$$

Macierz separacji $\mathbf{S} = [s_{ij}]_{n\times mn}$, $s_{ij} \in \{0,1\}$, oraz wektor $\mathbf{d} = [d_i]_{n\times 1}$, $d_i = 1$, blokują możliwość jednoczesnego umieszczenia na węźle $i = 1, \ldots, n$ wskazanej pary maszyn wirtualnych. W przypadku pary maszyn $p \neq r$ wyraża to następujące ograniczenie:

$$w_k + w_l \leq 1,\ k = (p-1)n+1, \ldots, (p-1)n+n,\ k = (r-1)n+1, \ldots, (r-1)n+n. \tag{3}$$

Konieczność separacji dotyczy maszyn wirtualnych oraz ich replik utrzymywanych przez podsystem zapewniania wysokiej dostępności usług, partycji baz danych oraz wybranych komponentów wymagających odseparowania ze względów bezpieczeństwa. Macierz $\mathbf{S}$ stanowi parametr zadania i może być określana przez administratora.

Ograniczenia postaci:

$$\underline{\mathbf{w}} \leq \mathbf{w} \leq \overline{\mathbf{w}} \tag{4}$$

pozwalają wymusić grupowanie maszyn wirtualnych na wskazanych maszynach fizycznych, przy czym $0 \leq \underline{w}_j \leq \overline{w}_j \leq 1$. Grupowaniu podlegać mogą maszyny tworzące grupę DMZ lub maszyny odpowiedzialne za filtrację pakietów pomiędzy sieciami VLAN. Wektor $\underline{\mathbf{w}}$ stanowi parametr zadania i może być określany przez administratora.

Struktura przyjętego modelu systemu pozwala wykorzystywać własność całkowitej unimodularności macierzy ograniczeń[1]. Jeżeli macierz ograniczeń jest unimodularna, całkowitoliczbowe (binarne) rozwiązanie $\hat{\mathbf{w}}$ problemu może być efektywnie obliczone bez konieczności wykorzystywania algorytmów podziału i ograniczeń lub w pierwszym kroku tego algorytmu. (W rozważanym przypadku, ze względu na strukturę macierzy separacji, warunek dostateczny unimodularności może nie być spełniony.)

2.2 Problem przydziału zasobów

Rozwiązanie $\hat{\mathbf{w}}$ problemu $\mathscr{P}$ określa następującą macierz lokalizacji (lub alokacji) maszyn wirtualnych:

$$\mathbf{H} = \mathbf{H}(\hat{\mathbf{w}}), \tag{5}$$

[1] Macierz jest całkowicie unimodularna, jeśli wyznacznik dowolnej jej podmacierzy kwadratowej (każdy minor macierzy) ma wartość 0, −1 lub 1.

gdzie $\mathbf{H} = [h_{ij}]_{n\times m},\ h_{ij} \in \{0,1\}$. Wiersze macierzy spełniają warunek:

$$\mathbf{H}[i,(1:m)] = \mathbf{A}[(i-1)n+1:in,(m+1:m+n)]^T\,\hat{\mathbf{w}}[(i-1)n+1:in],\ i = 1,\dots,n, \tag{6}$$

przy czym $\mathbf{H}(\hat{\mathbf{w}})$ można skonstruować przekształcając wektor $\hat{\mathbf{w}}$ z postaci wektora o wymiarze $mn\times 1$ do postaci macierzy o wymiarze $n\times m$.

Dysponując macierzą lokalizacji maszyn wirtualnych $\mathbf{H}(\hat{\mathbf{w}})$ w kolejnym kroku sformułować można następujące zadanie optymalizacji przydziału zasobów fizycznych klastra:

$$\mathscr{R}[(U_1,\dots,U_m),\mathbf{b},\underline{\mathbf{x}},\overline{\mathbf{x}}]:$$
$$\left|\begin{array}{ll} \text{zmaksymalizuj} & \displaystyle\sum_{s=1}^{m} U_s(\mathbf{x}_s) \\ \text{przy ograniczeniach} & \begin{pmatrix} \mathbf{H}(\hat{\mathbf{w}}) & & & \\ & \mathbf{H}(\hat{\mathbf{w}}) & & \\ & & \mathbf{H}(\hat{\mathbf{w}}) & \\ & & & \mathbf{H}(\hat{\mathbf{w}}) \end{pmatrix}\mathbf{x} \le \mathbf{b},\ \underline{\mathbf{x}} \le \mathbf{x} \le \overline{\mathbf{x}}. \end{array}\right.$$

Rozwiązaniem powyższego zadania jest wektor $\hat{\mathbf{x}} = [x_s]_{4m\times 1}$ przydziału zasobów maszyny fizycznej. Określa ono jaką część rdzeni CPU, pamięci RAM, dysku oraz pojemności łącza powinna otrzymać każda z maszyn wirtualnych alokowanych na maszynie fizycznej, aby zmaksymalizowana została sumaryczna użyteczność zasobów klastra. Użyteczność zasobów oceniana jest z punktu widzenia profilu użyteczności maszyn wirtualnych. W rozważanym przypadku przyjęto, że każda maszyna wirtualna $s = 1,\dots,m$ charakteryzuje się następującym modelem preferencji dotyczącym przydzielanych jej zasobów:

$$U_s(\mathbf{x}_s) = \begin{pmatrix} \gamma^s_{\mathtt{CPU}} \\ \gamma^s_{\mathtt{RAM}} \\ \gamma^s_{\mathtt{HDD}} \\ \gamma^s_{\mathtt{NET}} \end{pmatrix}^T \begin{pmatrix} \log(1+x_{1+4(s-1)}) \\ \log(1+x_{2+4(s-1)}) \\ \log(1+x_{3+4(s-1)}) \\ \log(1+x_{4+4(s-1)}) \end{pmatrix},\ s = 1,\dots,m. \tag{7}$$

Współczynniki $\gamma^s_{\mathtt{CPU}}$, $\gamma^s_{\mathtt{RAM}}$, $\gamma^s_{\mathtt{HDD}}$, $\gamma^s_{\mathtt{NET}}$ określają krańcową użyteczność zasobów fizycznych, jaką maszyna wirtualna uzyskuje realizując usługi systemowe. Im wyższa wartość współczynnika, tym cenniejszy jest zasób fizyczny z punktu widzenia usługi. Wartości współczynników są parametrem zadania i zostały ustalone eksperymentalnie w ramach przeprowadzonej diagnostyki systemu.

Ponieważ macierz ograniczeń wyznacza wypukły zbiór rozwiązań dopuszczalnych, a funkcje celu są ściśle wklęsłe, w zbiorze $\mathbb{R}^m$ istnieje jednoznaczne rozwiązanie sformułowanego powyżej problemu optymalizacji. Rozwiązanie to musi zostać zrzutowane na zbiór $\mathbb{Z}^m$ w celu uzyskania rozwiązania całkowitoliczbowego.

References

1. Author, F.: Article title. Journal **2**(5), 99–110 (2016)
2. Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016). https://doi.org/10.10007/1234567890
3. Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999)
4. Author, A.-B.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010)
5. LNCS Homepage, http://www.springer.com/lncs. Last accessed 4 Oct 2017