УДК 681.3

С. В. Юхимчук, д. т. н., проф.; Д. А. Білоус

ВИЗНАЧЕННЯ МАКСИМАЛЬНОГО ПОТОКУ В ДЕЦЕНТРАЛІЗОВАНИХ МЕРЕЖАХ

У роботі запропоновано методику обчислення максимального потоку в децентралізованих мережах, що базується на використанні адаптованого алгоритму Голдберга – Рао. Наведено основні припущення та вимоги щодо зведення задачі пошуку максимального потоку у децентралізованій мережі до задачі в орієнтованій мережі з єдиними джерелом та стоком. Для визначення максимального потоку запропоновано ітераційний алгоритм із використанням бінарних функцій довжин дуг.

***Ключові слова**: децентралізовані мережі, алгоритм Голдберга – Рао, максимальний потік.*

Вступ

Одноранговою, пірінговою або децентралізованою мережею (ДМ) називають обчислювальну мережу, що базується на рівності складових [1]. У таких мережах відсутні виділені сервери, а кожний вузол (peer) розглядається і як клієнт, і як сервер. На відміну від архітектури клієнт-сервер, така організація дозволяє мережі зберігати функціональність за будь-якої кількості та будь-якої комбінації вузлів. ДМ отримали застосування для організації файлообмінних мереж (E-Donkey, BitTorrent), інтернет-телебачення, інтернет-радіо, мультимедійних мереж тощо.

З огляду на стрімке зростання кількості ДМ у багатьох сферах, актуальною науковою проблемою є визначення ефективності функціонування ДМ певної структури та конфігурації. Визначення ефективності ДМ може дозволити вже на етапі проектування або під час експлуатації визначити «слабкі» місця ДМ та шляхи їх усунення, а також відповідність цієї структури поставленим задачам [1, 2].

Однією з найефективніших технологій використання ДМ сьогодні є виконання розподілених обчислень [1], що дозволяють за порівняно невеликий час виконувати величезний обсяг обчислень, який навіть при використанні суперкомп’ютерів потребував би, в залежності від складності задачі десятки або сотні років [2]. Така продуктивність досягається завдяки тому, що глобальна задача розбивається на велику кількість блоків, які одночасно виконуються сотнями тисяч комп’ютерів, що беруть участь у проекті. У таких випадках актуальною задачею постає визначення ефективності функціонування ДМ із врахуванням її конфігурації та характеристик її складових. Нині не існує загальновизнаного переліку показників ефективності функціонування обчислювальної мережі (ОМ), що пояснюється, в першу чергу, існуванням різних поглядів на основну мету її створення. До показників ефективності ОМ найчастіше відносять [1, 2, 3, 4]:

– пропускну спроможність;
– час доставки повідомлень;
– імовірність вчасної доставки повідомлень;
– надійність та живучість;
– рентабельність;
– вартість;
– ефективність керування.

З огляду на необхідність своєчасного обміну інформацією під час виконання розподілених обчислень, одним із головних критеріїв ефективності функціонування ДМ як ОМ можна вважати значення максимального потоку інформації, що може циркулювати у ДМ [2]. Тому серед задач визначення показників продуктивності та ефективності функціонування ДМ однією із найактуальніших є задача визначення максимального потоку у ДМ вже на етапі проектування або під час функціонування. Значення максимального потоку залежить від пропускної спроможності каналів зв'язку, апаратних ресурсів вузлів тощо, тому може розглядатися як комплексний показник ефективності функціонування ДМ. Крім того знаходження величини максимального потоку може дозволити визначити ефективність і доцільність використання каналів зв'язку та вузлів, що, в свою чергу, може розглядатися як ефективність даної структури та конфігурації ДМ.

Задача обчислення максимального потоку в мережі є класичною задачею оптимізації, алгоритми рішення якої розроблялися протягом останніх 50 – 60 років. Методи рішення застосовуються у транспортних, комунікаційних, електричних мережах, при моделюванні різноманітних фізичних та хімічних процесів тощо. Початковим підходом до розв'язку задачі знаходження максимального потоку було застосування симплекс-методу лінійного програмування. У 1956 році Фордом та Фалкерсоном для рішення цієї задачі було запропоновано використання орієнтованої мережі та ітераційного псевдополіноміального алгоритму, що дозволило поліпшити оцінку алгоритму знаходження максимального потоку з $O(n^2 mU)$ (де $n = |V|$ – кількість вершин, $m = |E|$ – кількість ребер, $U = \max(C_{ij})$ – максимальне значення пропускної спроможності) до $O(nm \cdot U)$. У цьому випадку $O(g(n))$ є верхньою асимптотичною оцінкою складності алгоритму і визначає швидкість його виконання в найгіршому випадку. Час виконання алгоритму $T(n)$ має оцінку $O(g(n))$, якщо $\exists c > 0, n_0 > 0 : T(n) \le cg(n), \forall n > n_0$ [5]. Подальші дослідження дозволили ще більше знизити оцінку алгоритму знаходження максимального потоку. Хронологія основних досягнень в рішенні задачі про максимальний потік [6] наведена у таблиці.

Таблиця

Хронологічна таблиця досягнень в рішенні задачі про максимальний потік

Рік	Автор	Оцінка алгоритму
1951	Данциг	$O(n^2 mU)$
1956	Форд , Фалкерсон	$O(nm^2 \cdot U)$
1970	Едмондс, Карп	$O(nm^2)$
1970	Дініц	$O(n^2 m)$
1972	Едмондс, Карп	$O(m^2 \cdot \log U)$
1973	Дініц, Габоу	$O(nm \cdot \log U)$
1974	Карзанов	$O(n^3)$
1978	Малхотрі, Кумар, Махешварі	$O(n^3)$
1978	Галіл, Намад	$O(nm \cdot \log^2 n)$
1980	Слейтор, Тар'ян	$O(nm \cdot \log n)$
1986	Голдберг, Тар'ян	$O(nm \cdot \log \frac{n^2}{m})$
1989	Черіян, Хагеруп, Мелхорн	$O(nm + n^2 \log^2 n)$
1992	Кінг, Рао, Тар'ян	$O(nm + n^{2+e})$
1997	Голдберг, Рао	$O(\min\left\{m^{\frac{2}{3}}, n^{\frac{1}{2}}\right\} m \cdot \log\left(\frac{n^2}{m}\right) \log U)$

У 1997 році Голдебергом та Рао був запропонований алгоритм, що надавав дугам неодиничну довжину. Сьогодні саме цей алгоритм є найбільш сучасним та ефективним, його асимптотична оцінка перевершила $O(nm)$. Переважна більшість сучасних досліджень спрямована на вивчення та поліпшення саме алгоритму Голдберга – Рао [6, 7].

З огляду на те, що структура орієнтованої на виконання розподілених обчислень ДМ, не втрачаючи зв'язності, може постійно змінюватись через відключення або випадкові відмови вузлів, існує необхідність у визначенні можливого значення максимального потоку для поточної конфігурації ДМ. **Метою** нашої роботи є розробка алгоритму визначення максимального потоку, коли, на відміну від класичних алгоритмів Голдберга – Рао та Форда – Фалкерсона, самі вузли мають обмежену пропускну спроможність.

Визначення максимального потоку у децентралізованій мережі

Математична модель децентралізованої мережі

В якості математичної моделі ДМ на рівні морфологічного опису пропонується використовувати орієнтований зважений граф $G(V, E)$, вершини якого V відповідають вузлам мережі, а ребра E – лініям зв'язку.

Кожній вершині $v_i \in V$ ставиться у відповідність такий набір показників і відповідних їм значень [3, 4]:

– пропускна спроможність $c(v_i)$;

– імовірність відмови вузла $q(v_i)$;

– обсяг буферної пам'яті вузла $r(v_i)$;

– вартість вузла $s(v_i)$;

– алгоритм обслуговування замовлень U_i.

Кожному ребру $e_j \in E$, що з'єднує вершини v та w, – такий набір показників [3, 4]:

– довжина ліній $l(e_j)$;

– імовірність відмови $q(k)$ каналу k лінії e_j (в загальному випадку відмови каналів є незалежними);

– ємність лінії $t(e_j)$;

– пропускна спроможність $c(e_j)$;

– імовірність спотворення $\xi(k)$ повідомлень завадами при передачі по каналу k лінії;

– вартість лінії $s(e_j)$.

У моделі передбачається незалежність відмови вузлів та ліній ДМ і вважається, що відмови є наслідком лише фізичної надійності елементів і не залежать від законів та ступеню їх використання.

Запропонована модель описує лише первинну ДМ як мережу ліній зв'язку. Оскільки вторинна мережа будується на базі каналів первинної за допомогою їх кросування на певних вузлах, модель первинної мережі необхідно доповнити відповідним планом кросування із врахуванням більш складних залежностей $\{q(V);V\}$. Для знаходження максимального потоку у мережі необхідною і достатньою є первинна модель [3], тому що вона в повній мірі враховує всі необхідні характеристики компонентів ДМ.

Зведення задачі пошуку максимального потоку у децентралізованій мережі до задачі в орієнтованій мережі з єдиним джерелом та єдиним стоком

З метою застосування алгоритмів знаходження максимального потоку для випадку ДМ потрібно звести дану задачу до задачі про знаходження максимального потоку в орієнтованій мережі з єдиним джерелом та єдиним стоком. Розглянемо властивості ДМ, які є відмінними від класичних умов:

1. Наявність декількох джерел та стоків.

У ДМ, що орієнтована на виконання розподілених обчислень, множину вузлів можна розділити на 3 групи: S – множина джерел, T – множина стоків, R – множина проміжних вузлів. Потік може бути спрямований з будь-якого джерела в будь-якій стік. У такому випадку графову модель ДМ слід модифікувати шляхом додаванням нової вершини S^* – суперджерела – та з’єднанням її з усіма вузлами множини S та додаванням нової вершини T^* – суперстоку – та з’єднанням її з усіма вузлами множини T. Як значення пропускних спроможностей кожної з нових дуг будемо вважати ∞. У такому разі пошук максимального потоку з вузлів S множини у вузли множини T є еквівалентним пошуку максимального потоку з S^* до T^* [6].

2. Обмеженість пропускних спроможностей вузлів

У ДМ можливий випадок існування пропускних спроможностей не лише у каналів зв’язку але й у вузлів, що пояснюється серед іншого обмеженням апаратних ресурсів вузлів. У такому разі слід кожну вершину $v_i \in V$, що має пропускну спроможність $c(v_i)$, замінити двома вершинами $v_i^{'}$ та $v_i^{''}$ та з’єднати їх дугою, пропускну спроможність якої прийняти за $c(v_i)$.

3. Випадок неорієнтованої або змішаної ДМ.

У ДМ можливий випадок існування неорієнтованих ребер, тобто каналів зв’язку, пропускна спроможність яких не залежить від напрямку обміну інформацією. У такому разі потік по неорієнтованій дузі (v,w) з пропускною спроможністю $c(v,w)$ повинен задовольняти вимоги:

$$\begin{cases} f(v,w) \le c(v,w), \\ f(w,v) \le c(v,w), \\ f(v,w) \cdot f(w,v) = 0. \end{cases}$$

Заміна неорієнтованої дуги двома протилежно спрямованими орієнтованими дугами з однаковими пропускними спроможностями $c(v,w)$ зводить задачу до еквівалентної задачі на орієнтованому графі [7]. Тобто у випадку неорієнтованого або змішаного графу ДМ потрібно кожну з неорієнтованих дуг замінити парою орієнтованих.

Алгоритм визначення максимального потоку у децентралізованій мережі

Для довільного випадку ДМ кількість дуг $m = |E|$ значно перевищує кількість вузлів $n = |V|$: $m >> n$, а в граничному випадку $m = n^2$. Тому найбільш вдалим буде застосування алгоритму Голдберга – Рао, оцінка якого в найгіршому випадку складе $O(n^{2\frac{1}{2}} \cdot \log U)$, що краще за оцінку алгоритму Карзанова ($O(n^3)$) та обох модифікацій алгоритму Тар’яна ($O(n^3 \cdot \log n)$ та $O(n^3)$) [5].

Нехай $G(V,E)$ – графова модель ДМ. При цьому підкреслимо, що структура ДМ відрізняється від структури мережі, яка використовується в класичному алгоритмі Голдберга – Рао. Тому до алгоритму слід ввести попередній етап, який дозволить звести існуючу задачу до задачі пошуку максимального потоку в класичній орієнтованій мережі. Створення такого етапу реалізується за допомогою наступних кроків:

1. Додамо до моделі суперджерело S^* та суперстік T^*.
2. Замінимо кожен вузол $v_i \in V$, що має пропускну спроможність $c(v_i)$, парою вузлів $v_i^{'}$ та $v_i^{''}$, з’єднавши їх дугою з пропускною спроможністю $c(v_i)$.
3. Замінимо кожну неорієнтовану дугу $e_j \in E$, що має пропускну спроможність $c(e_j)$, парою протилежно спрямованих дуг з однаковими пропускними спроможностями $c(e_j)$.

4. У результаті отримаємо модифіковану графову модель ДМ $G'(V', E')$ з джерелом S^* та стоком T^*.

Нехай функція $c(e): E' \to \{1,...,C\}$ визначає пропускні спроможності каналів зв'язку. Потоком, відповідно, назвемо функцію $f(e): E' \to \{1,...,C\}$, для якої виконується умова $f(e) \le c(e)$ та принцип збереження потоку. Залишковою пропускною спроможністю є $c_f(i,j) = f(i,j) + c(j,i) - c(i,j)$. Дуги, що мають додатну c_f, складають множину E_f. Залишковий потік є різницею між деяким оптимальним потоком у ДМ f^* та поточним f^R: $\left|f^*(e) - f^R(e)\right|$. Тупиковим потоком назвемо функцію, при якій у графі E_f не існує шляху з S^* до T^*.

Функцією довжини дуги назвемо функцію $l: E' \to R^+$. Функцію $d: V' \to R^+$ назвемо міткою дистанції відносно l, якщо $d(t) = 0$ та для всіх дуг $e \in E'$ залишкова вартість невід'ємна. Залишкова вартість визначається як $l_d(i,j) = l(i,j) + d(j) - d(i)$. Залишкова вартість на дугах шляху з S^* до T^* може лише зростати. Позначимо $d_l(i)$ дистанцію вершини i до t відносно l. Крім того для будь-якого d виконується $\forall d, d \le d_l$ [6,8].

За допомогою функції довжини дуги можна отримати верхню границю залишкового потоку. Ребро з E_f можна розглядати як канал шириною $c(e)$, довжиною $l(e)$ та площею $Ar(e) = c(e)l(e)$. Площею мережі Ar_{Ef} буде сума площ всіх ребер з додатною залишковою пропускною спроможністю [6]. Нехай кожна одиниця залишкового потоку потребує $d(s)$ одиниць площі, тоді верхньою границею величини залишкового потоку буде $\dfrac{Ar_{Ef}}{d(s)}$. Назвемо ребро (i,j) із E_f допустимим, якщо $d(i) > d(j)$ або $d(i) = d(j)$ при $l(,i,j) = 0$. З допустимих ребер формується множина $A(f,d,l)$ та складається граф $G_A(V', A(f,d,l))$.

Алгоритм Голдберга – Рао базується на бінарній функції довжини дуги l', яка дорівнює 0 на ребрах з «великою» залишковою пропускною спроможністю та 1 – на решті [8]. Це дозволяє зменшити Ar_{Ef} і отримати кращу оцінку для залишкового потоку в порівнянні з одиничною функцією довжини. Визначення «великих» залишкових пропускних спроможностей залежить від поточної величини залишкового потоку.

На кожній ітерації алгоритму визначається поточне значення залишкового потоку F та відповідне значення ΔF – величини, на яку потрібно збільшити значення F на даному кроці алгоритму. Бінарна функція довжини $l(e)$ дорівнює 0, якщо $c_f(e) \ge 3\Delta F$, і 1 – в протилежному випадку. Довжина деяких ребер змінюється після нарощування потоку в мережі за допомогою тупикового потоку [8].

На початку кожної ітерації з графу G_A видаляються всі компоненти сильної зв'язності, утворені ребрами нульової довжини, і збільшується потік в отриманому ациклічному графі, для якого або знаходиться тупиковий потік, або потік величини ΔF.

На відміну від класичного алгоритму Голдберга – Рао, у нашій роботі пропонується використовувати для визначення ΔF замість бінарного пошуку алгоритм знаходження максимального потоку через розріз мережі, описаний в [6]. Така заміна дозволить підвищити швидкодію алгоритму за рахунок зменшення числа обчислювальних операцій, пов'язаних із здійсненням відповідного перебору.

Загалом роботу алгоритму на i-ій ітерації можна представити у вигляді такої послідовності дій:

Поки $F \ge 1$

1. Оновлюється параметр ΔF , бінарна функція довжини $l^{'}$ та мітки дистанцій d .
2. Вилучаються компоненти сильної зв’язності, утворені ребрами нульової довжини. Вилучення здійснюється лише в межах поточної ітерації, на наступній – ребра поновлюються.
3. Визначається граф $G_A(V^{'}, A(f, d, l))$ допустимих ребер.
4. Визначається потік у G_A , який або є тупиковим, або дорівнює ΔF .
5. Нарощується поточний потік потоком, що був знайдений на попередньому кроці.
6. Оновлюється F .

Кінець.

Знаходження тупикового потоку здійснюється будь-яким із існуючих методів. Якщо величина знайденого потоку $H > \Delta F$, надлишок усувається [6].

Запропоновані модифікації алгоритму Годберга – Рао полягають у його адаптації до специфіки структури ДМ у порівняні з класичною мережею із єдиним джерелом та єдиним стоком та у оптимізації визначення зміни F через розріз мережі замість бінарного пошуку. Запропоновані модифікації не впливають на збіжність, отже, збіжність модифікованого алгоритму може бути доведена аналогічно до доведення збіжності оригінального алгоритму Голдберга – Рао, що детально розглянуте в [8].

Висновки

У цій роботі розглянуто метод визначення максимального потоку в ДМ. Запропоновано алгоритм розв’язку задачі визначення максимального потоку ДМ, що базується на алгоритмі Голдберга – Рао. Асимптотична оцінка запропонованого алгоритму становить $O(n^{2\frac{1}{2}} \cdot \log U)$, що задовольняє вимоги поставленої мети – розробити алгоритм із оцінкою, меншою за $O(n^3)$.

Удосконалення алгоритму можливе за рахунок використання паралельних обчислень, що може дозволити знизити оцінку роботи алгоритму до $O(n)$.

Надалі планується розробити пакет програмних засобів, що здійснюватиме моделювання поведінки ДМ та дозволятиме визначити числові значення максимального потоку ДМ. Керуючись проміжними значеннями роботи алгоритму, можливе визначення «слабких» місць ДМ, випадків неефективного використання ресурсів вузлів або каналів тощо.

СПИСОК ЛІТЕРАТУРИ

1.Cisco Systems. Руководство по технологиям объединенных сетей. – М: Вильямс, 2004. – 1039 с.

2.Менаске Д., Алмейда В. Производительность web-служб. Анализ, оценка и планирование. – М: Diasoft, 2005. – 467 с.

3.Ченцов В.М. Системы распределения информации. Синтез структуры и управления. – М: Связь, 1980. – 144 с.

4.Білоус Д.А., Юхимчук С.В. Комплексний критерій оцінювання ефективності функціонування обчислювальних мереж // Труды восьмой международной научно-практической конференции «Современные информационные и электронные технологии, Одеса, 2007. – С. 117.

5.Кормен Т. Алгоритмы: построение и анализ. – М: МЦНМО, 2002. – 957 с.

6.Hudson I. The Maximum Flow Problem. – Denison University, 2004. – 19 с.

7.Касьянов В.Н., Евстигнеев В.А. Графы в программировании: обработка, визуализация и применение. – Санкт-Петербург: «БХВ-Петербург», 2003. – 1104 с.

8.Papp D. The Goldberg-Rao Algorithm for the Maximum Flow Problem. – COS, 2006. – 5 p.

Юхимчук Сергій Васильович – завідувач кафедри;

Білоус Дмитро Анатолійович – магістрант кафедри.

Кафедра інтелектуальних систем, Вінницький національний технічний університет