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

ОПРЕДЕЛЕНИЕ МАКСИМАЛЬНОГО ПОТОКА В ДЕЦЕНТРАЛИЗОВАННЫХ СЕТЯХ

В работе предложена методика определения максимального потока в децентрализованных сетях, которая основывается на использовании адаптированного алгоритма Голдберга – Рао. Приведены основные предположения и требования по приведению задачи поиска максимального потока в децентрализованной сети к задаче в ориентированной сети с единственным источником и стоком. Для определения максимального потока предложен итерационный алгоритм с использованием бинарных функций длин дуг.

Ключевые слова: децентрализованные сети, алгоритм Голдберга – Рао, максимальный поток.

Введение

Одноранговой, пиринговой или децентрализованной сетью (ДС) называют вычислительную сеть, которая базируется на равенстве составляющих [1]. В таких сетях отсутствуют выделенные серверы, а каждый узел (peer) рассматривается и как клиент, и как сервер. В отличие от архитектуры клиент-сервер, такая организация позволяет сети сохранять функциональность при любом количестве и любой комбинации узлов. ДС получили применение для организации файлообменных сетей (E-Donkey, BitTorrent), интернет-телевидения, интернет-радио, мультимедийных сетей и т.д.

Учитывая стремительный рост количества ДС во многих сферах, актуальной научной проблемой являются определение эффективности функционирования ДС определенной структуры и конфигурации. Определение эффективности ДС может позволить уже на этапе проектирования или во время эксплуатации определить «слабые» места ДС и пути их устранения, а также соответствие данной структуры поставленным задачам [1, 2].

Одной из наиболее эффективных технологий использования ДС сегодня является выполнение распределенных вычислений [1], которые позволяют за сравнительно небольшое время выполнять огромный объем вычислений, который даже при использовании суперкомпьютеров потребовал бы в зависимости от сложности задачи десятки или сотни лет [2]. Такая производительность достигается благодаря тому, что глобальная задача разбивается на большое количество блоков, которые одновременно выполняются сотнями тысяч компьютеров, принимающих участие в проекте. В таких случаях актуальной задачей является определение эффективности функционирования ДС с учетом ее конфигурации и характеристик ее составляющих. Ныне не существует общепризнанного перечня показателей эффективности функционирования вычислительной сети (ВС), что объясняется, в первую очередь, существованием разных взглядов на основную цель ее создания. К показателям эффективности ОМ чаще других относятся [1, 2, 3, 4]:

- пропускная способность;
- время доставки сообщений;
- вероятность своевременной доставки сообщений;
- надежность и живучесть;
- рентабельность;
- стоимость;
- эффективность управления.

Учитывая необходимость своевременного обмена информацией во время выполнения распределенных вычислений, одним из главных критериев эффективности функционирования ДС как ВС можно считать значение максимального потока информации, который может циркулировать в ДС [2]. Поэтому среди задач определения показателей

производительности и эффективности функционирования ДС одной из наиболее актуальных является задача определения максимального потока в ДС уже на этапе проектирования или во время функционирования. Значение максимального потока зависит от пропускной способности каналов связи, аппаратных ресурсов узлов и т.п., поэтому может рассматриваться как комплексный показатель эффективности функционирования ДС. Кроме того нахождение величины максимального потока может позволить определить эффективность и целесообразность использования каналов связи и узлов, что, в свою очередь, может рассматриваться как эффективность данной структуры и конфигурации ДС.

Задача вычисления максимального потока в сети является классической задачей оптимизации, алгоритмы решения которой разрабатывались на протяжении последних 50 – 60 лет. Методы решения применяются в транспортных, коммуникационных, электрических сетях, при моделировании разнообразных физических и химических процессов и т.п.

Изначальным подходом к решению задачи нахождения максимального потока было применение симплекс-метода линейного программирования. В 1956 году Фордом и Фалкерсоном для решения данной задачи было предложено использование ориентированной сети и итерационного псевдополиномиального алгоритма, что позволило улучшить оценку алгоритма нахождения максимального потока с $O(n^2mU)$ (где $n = |V|$ – количество вершин, $m = |E|$ – количество ребер, $U = \max(C_{ij})$ – максимальное значение пропускной способности) до $O(nm \cdot U)$. В данном случае $O(g(n))$ является верхней асимптотической оценкой сложности алгоритма и определяет скорость его выполнения в худшем случае. Время выполнения алгоритма $T(n)$ имеет оценку $O(g(n))$, если $\exists c > 0, n_0 > 0 : T(n) \leq cg(n), \forall n > n_0$ [5]. Следующие исследования позволили еще больше снизить оценку алгоритма нахождения максимального потока. Хронология основных достижений решения задачи о максимальном потоке [6] приведена в таблице.

Таблица

Хронологическая таблица достижений решения задачи о максимальном потоке

Год	Автор	Оценка алгоритма
1951	Данциг	$O(n^2mU)$
1956	Форд, Фалкерсон	$O(nm^2 \cdot U)$
1970	Эдмондс, Карп	$O(nm^2)$
1970	Диниц	$O(n^2m)$
1972	Эдмондс, Карп	$O(m^2 \cdot \log U)$
1973	Диниц, Габоу	$O(nm \cdot \log U)$
1974	Карзанов	$O(n^3)$
1978	Малхотри, Кумар, Махешвари	$O(n^3)$
1978	Галил, Намад	$O(nm \cdot \log^2 n)$
1980	Слейтор, Тарьян	$O(nm \cdot \log n)$
1986	Голдберг, Тарьян	$O(nm \cdot \log \frac{n^2}{m})$
1989	Чериян, Хагеруп, Мелхорн	$O(nm + n^2 \log^2 n)$
1992	Кинг, Рао, Тарьян	$O(nm + n^{2+e})$
1997	Голдберг, Рао	$O(\min\left\{m^{\frac{2}{3}}, n^{\frac{1}{2}}\right\} m \cdot \log\left(\frac{n^2}{m}\right) \log U)$

В 1997 году Голдберг и Рао предложили алгоритм, приписывающий дугам неединичную длину. Сегодня именно этот алгоритм является наиболее современным и эффективным, его асимптотическая оценка превзошла $O(nm)$. Подавляющее большинство современных исследований направлены на изучение и улучшение именно алгоритма Голдберга – Рао [6, 7].

Учитывая то, что структура ориентированной на выполнение распределенных вычислений ДС, не теряя связности, может постоянно изменяться из-за отключения или случайных отказов узлов, существует необходимость в определении возможного значения максимального потока для текущей конфигурации ДС. **Целью** данной работы является разработка алгоритма определения максимального потока в случае, когда, в отличие от классических алгоритмов Голдберга – Рао и Форда – Фалкерсона, узлы имеют ограниченную пропускную способность.

Определение максимального потока в децентрализованной сети

Математическая модель децентрализованной сети

В качестве математической модели ДС на уровне морфологического описания предлагается использовать ориентированный взвешенный граф $G(V, E)$, вершины которого V отвечают узлам сети, а ребра E – линиям связи.

Каждой вершине $v_i \in V$ ставится в соответствие такой набор показателей и соответствующих им значений [3,4]:

- пропускная способность $c(v_i)$;
- вероятность отказа узла $q(v_i)$;
- объем буферной памяти узла $r(v_i)$;
- стоимость узла $s(v_i)$;
- алгоритм обслуживания заказов U_i .

Каждому ребру $e_j \in E$, соединяющему вершины v_i и w – следующий набор показателей [3, 4]:

- длина линий $l(e_j)$;
- вероятность отказа $q(k)$ канала k линии e_j (в общем случае отказа каналов являются независимыми);
- емкость линии $t(e_j)$;
- пропускная способность $c(e_j)$;
- вероятность искажения $\xi(k)$ сообщений помехами при передаче по каналу k линии;
- стоимость линии $s(e_j)$.

В модели предполагается независимость отказа узлов и линий ДС и считается, что отказы являются следствием лишь физической надежности элементов и не зависят от законов и степени их использования.

Предложенная модель описывает лишь первичную ДС как сеть линий связи. Так как вторичная сеть строится на базе каналов первичной с помощью их кроссировки на определенных узлах, модель первичной сети необходимо дополнить соответствующим планом кроссировки с учетом более сложных зависимостей $\{q(V); V\}$. Для нахождения максимального потока в сети необходимой и достаточной является первичная модель [3], поскольку она в полной мере учитывает все необходимые характеристики компонентов ДС.

Сведение задачи поиска максимального потока в децентрализованной сети к задаче в ориентированной сети с единственным источником и стоком

С целью применения алгоритмов нахождения максимального потока для случая ДС нужно

свести данную задачу к задаче о нахождении максимального потока в ориентированной сети с единственным источником и единственным стоком. Рассмотрим свойства ДС, которые отличаются от классических условий:

1. Наличие нескольких источников и стоков.

В ДС, ориентированной на выполнение распределенных вычислений, множество узлов можно разделить на 3 группы: S – множество источников, T – множество стоков, R – множество промежуточных узлов. Поток может быть направлен из любого источника в любой сток. В таком случае графовую модель ДС следует модифицировать путем добавления новой вершины S^* – суперисточника – и соединения ее со всеми узлами множества S и добавления новой вершины T^* – суперстока – и соединения ее со всеми узлами множества T . В качестве значений пропускных способностей каждой из новых дуг будем считать ∞ . В таком случае поиск максимального потока из узлов S множества в узлы множества T является эквивалентным поиску максимального потока из S^* в T^* [6].

2. Ограниченность пропускных способностей узлов

В ДС возможен случай существования пропускных способностей не только у каналов связи, но и у узлов, который объясняется ограничением аппаратных ресурсов узлов. В таком случае следует каждую вершину $v_i \in V$, которая имеет пропускную способность $c(v_i)$, заменить двумя вершинами v_i' и v_i'' и соединить их дугой, пропускную способность которой принять за $c(v_i)$.

3. Случай неориентированной или смешанной ДС.

В ДС возможен случай существования неориентированных ребер, то есть каналов связи, пропускная способность которых не зависит от направления обмена информацией. В таком случае поток по неориентированной дуге (v, w) с пропускной способностью $c(v, w)$ должен удовлетворять требованиям:

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

Замена неориентированной дуги двумя противоположно направленными ориентированными дугами с одинаковыми пропускными способностями $c(v, w)$ сводит задачу к эквивалентной задаче на ориентированном графе [7]. То есть в случае неориентированного или смешанного графа ДС нужно каждую из неориентированных дуг заменить парой ориентированных.

Алгоритм определения максимального потока в децентрализованной сети

Для произвольного случая ДС количество дуг $m = |E|$ значительно превышает количество узлов $n = |V|$: $m \gg n$, а в предельном случае $m = n^2$. Поэтому наиболее удачным будет применение алгоритма Голдберга – Рао, оценка которого в худшем случае составит $O(n^{2\frac{1}{2}} \cdot \log U)$, что лучше оценки алгоритма Карзанова ($O(n^3)$) и обеих модификаций алгоритма Тарьяна ($O(n^3 \cdot \log n)$ и $O(n^3)$) [5].

Пусть $G(V, E)$ – графовая модель ДС. При этом подчеркнем, что структура ДС отличается от структуры сети, которая используется в классическом алгоритме Голдберга – Рао. Поэтому к алгоритму следует добавить предварительный этап, который позволит свести существующую задачу к задаче поиска максимального потока в классической ориентированной сети. Создание такого этапа реализуется с помощью следующих шагов:

1. Добавим к модели суперисточник S^* и суперсток T^* .
2. Заменяем каждый узел $v_i \in V$, имеющий пропускную способность $c(v_i)$, парой узлов v_i' и v_i'' , соединив их дугой с пропускной способностью $c(v_i)$.
3. Заменяем каждую неориентированную дугу $e_j \in E$, имеющую пропускную способность $c(e_j)$, парой противоположно направленных дуг с одинаковыми пропускными способностями $c(e_j)$.
4. В результате получим модифицированную графовую модель ДМ $G'(V', E')$ с источником S^* и стоком T^* .

Пусть функция $c(e): E' \rightarrow \{1, \dots, C\}$ определяет пропускные способности каналов связи. Поток, соответственно, назовем функцию $f(e): E' \rightarrow \{1, \dots, C\}$, для которой выполняется условие $f(e) \leq c(e)$ и принцип сохранения потока. Остаточной пропускной способностью назовем $c_f(i, j) = f(i, j) + c(j, i) - c(i, j)$. Дуги, имеющие положительную c_f , составляют множество E_f . Остаточный поток является разностью между некоторым оптимальным потоком в ДС f^* и текущим $f^R: |f^*(e) - f^R(e)|$. Тупиковым потоком назовем функцию, при которой в графе E_f не существует пути из S^* в T^* .

Функцией длины дуги назовем функцию $l: E' \rightarrow R^+$. Функцию $d: V' \rightarrow R^+$ назовем меткой дистанции относительно l , если $d(t) = 0$ и для всех дуг $e \in E'$ остаточная стоимость неотрицательная. Остаточная стоимость определяется как $l_d(i, j) = l(i, j) + d(j) - d(i)$. Остаточная стоимость на дугах пути из S^* в T^* может лишь возрастать. Обозначим $d_l(i)$ дистанцию вершины i до t относительно l . Кроме того заметим, что для любого d выполняется $\forall d, d \leq d_l$ [6, 8].

С помощью функции длины дуги можно получить верхнюю границу остаточного потока. Ребро из E_f можно рассматривать как канал шириной $c(e)$, длиной $l(e)$ и площадью $Ar(e) = c(e)l(e)$. Площадью сети Ar_{E_f} будет сумма площадей всех ребер с положительной остаточной пропускной способностью [6]. Пусть каждая единица остаточного потока нуждается в $d(s)$ единицах площади, тогда верхней границей величины остаточного потока будет $\frac{Ar_{E_f}}{d(s)}$. Назовем ребро (i, j) из E_f допустимым, если $d(i) > d(j)$ или $d(i) = d(j)$ при $l(i, j) = 0$. Из допустимых ребер формируется множество $A(f, d, l)$ и составляется граф $G_A(V', A(f, d, l))$.

Алгоритм Голдберга – Рао базируется на бинарной функции длины дуги l' , которая равняется 0 на ребрах с «большой» остаточной пропускной способностью и 1 – на остальных [8]. Это позволяет уменьшить Ar_{E_f} и получить лучшую оценку для остаточного потока в сравнении с единичной функцией длины. Определение «больших» остаточных пропускных способностей зависит от текущей величины остаточного потока.

На каждой итерации алгоритма определяется текущее значение остаточного потока F и соответствующее значение ΔF – величины, на которую нужно увеличить значение F на данном шаге алгоритма. Бинарная функция длины $l(e)$ равняется 0, если $c_f(e) \geq 3\Delta F$, и 1 – в противоположном случае. Длина некоторых ребер изменяется после наращивания потока в сети с помощью тупикового потока [8].

В начале каждой итерации из графа G_A удаляются все компоненты сильной связности,

которые образованы ребрами нулевой длины, и увеличивается поток в полученном ациклическом графе, для которого или находится тупиковый поток, или поток величины ΔF .

В отличие от классического алгоритма Голдберга – Рао в данной работе предлагается использовать для определения ΔF вместо бинарного поиска алгоритм нахождения максимального потока через разрез сети, который описан в [6]. Такая замена разрешит повысить быстродействие алгоритма за счет уменьшения числа вычислительных операций, связанных с осуществлением соответствующего перебора.

В общем случае работу алгоритма на i -ой итерации можно представить в виде следующей последовательности действий:

Пока $F \geq 1$

1. Обновляется параметр ΔF , бинарная функция длины l' и метки дистанций d .
2. Изымаются компоненты сильной связности, образованные ребрами нулевой длины. Изъятия осуществляется лишь в пределах текущей итерации, на следующей ребра возобновляются.
3. Определяется граф $G_A(V', A(f, d, l))$ допустимых ребер.
4. Определяется поток в G_A , который или есть тупиковым, или равняется ΔF .
5. Нарастивается текущий поток потоком, который был найден на предыдущем шаге.
6. Обновляется F .

Конец.

Нахождение тупикового потока осуществляется любым из существующих методов. Если величина найденного потока $H > \Delta F$, излишек устраняется [6].

Предложенные модификации алгоритма Голдберга – Рао состоят в его адаптации к специфике структуры ДС в сравнении с классической сетью с единственным источником и единственным стоком и в оптимизации определения изменения F через разрез сети вместо бинарного поиска. Предложенные модификации не влияют на сходимость, а значит сходимость модифицированного алгоритма может быть доказана аналогично доказательству сходимости оригинального алгоритма Голдберга – Рао, детально рассмотренному в [8].

Выводы

В данной работе рассматривается метод определения максимального потока в ДС. Предложен алгоритм решения задачи определения максимального потока ДС, который базируется на алгоритме Голдберга – Рао. Асимптотическая оценка предложенного алгоритма составляет $O(n^{2\frac{1}{2}} \cdot \log U)$.

Усовершенствование алгоритма возможно за счет использования параллельных вычислений, что может позволить снизить оценку работы алгоритма до $O(n)$.

В дальнейшем планируется разработать пакет программных средств, который будет осуществлять моделирование поведения ДС и позволит определять числовые значения максимального потока ДС. Руководствуясь промежуточными значениями работы алгоритма возможно определение «слабых» мест ДС, случаев неэффективного использования ресурсов узлов или каналов и т.п.

СПИСОК ЛИТЕРАТУРЫ

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 с.

Юхимчук Сергей Васильевич – заведующий кафедрой;

Билоус Дмитрий Анатольевич – магистрант кафедры.

Кафедра интеллектуальных систем, Винницкий национальный технический университет