

Применение метода агрегации данных в задаче нахождения узлов с дополнительной функциональностью¹

Я.Р. Гринберг, И.И. Курочкин, А.В. Корх, Р.М. Алыгулиев, М.А. Гашимов

Аннотация. Решается задача нахождения в телекоммуникационных сетях связной подсети узлов с дополнительной функциональностью для организации двухуровневой маршрутизации. Выделение таких узлов производится при наличии у них определенных характеристик с помощью метода агрегации данных элементов сети. Производится сравнение этих результатов с аналогичным, полученным ранее методом минимальных разрезов на графе признаков.

Ключевые слова: Сеть, маршрутизация, суперпир, выделение суперпиров, функция разгруженности, агрегация данных, магистральная сеть.

Введение

Данная работа посвящена проблеме динамической маршрутизации потоков данных в телекоммуникационных сетях, которая может быть рассмотрена в рамках многопродуктовой задачи. Цель настоящего рассмотрения - максимизация суммарного потока в сети между парами корреспондирующих узлов, называемых полюсами сети, при ограниченных пропускных способностях каналов связи и последовательном во времени поступлении заявок. Для решения задачи были использованы алгоритмы последовательного заполнения сети потоками продуктов [2-4]. В их основу положен поиск пути наименьшей стоимости, причем стоимости ребер вычислялись в зависимости от принадлежности ребер минимальному разрезу той или иной пары полюсов. Такой подход предполагает большую вычислительную сложность предложенных алгоритмов, вследствие необходимости нахождения всех минимальных разрезов между

всеми парами полюсов. Вычислительная сложность минимально-разрезных алгоритмов была $O(|V||E|^2 P)$, где V - число узлов сети, E - число ребер графа, описывающего телекоммуникационную сеть, P - число пар корреспондирующих узлов (полюсов сети). Одним из вариантов уменьшения времени работы алгоритмов последовательного заполнения сети является декомпозиция задачи динамической маршрутизации и введение двухуровневой маршрутизации. Двухуровневая маршрутизация предполагает выделение в сети узлов с дополнительной функциональностью (называемых также суперпирами), которая подразумевает организацию локальной маршрутизации. Суперпиры образуют связную подсеть. Потоки продуктов идут от узла-источника к одному из суперпиров, по подсети суперпиров и оканчиваются на узле-адресате.

Предложенный в работе [1] способ выделения узлов с дополнительной функциональностью предполагал точное решение задачи бина-

¹ Работа выполнена при поддержке ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы» (Госконтракт № 11.519.11.4012) и РФФИ (грант № 12-07-00415-а; грант № 12-07-31227-мол-а.)

ризации (разделения на два класса) узлов телекоммуникационной сети, однако его основными недостатками были:

1. централизованность, т.е. необходимость получения одним узлом данных обо всей сети для разбиения множества узлов на суперпиры и остальные узлы;

2. требование знания «эталонных» узлов, т.е. типичных представителей класса суперпиров и обычных клиентских узлов;

3. необходимость использования настроечных параметров алгоритма, значение которых получаются из эмпирических соображений;

4. как следствие этого, невозможность прямого контроля числа находимых суперпиров.

Предлагаемый в данной работе метод позволяет преодолеть указанные недостатки.

Во второй части работы дан краткий обзор существующих на данный день алгоритмов работы с выделяемыми в сети суперпирами. В третьей части дана постановка задачи, определены особенности решаемой задачи и дано описание алгоритма. В четвертой части представлено подробное описание экспериментов, пятый раздел посвящен их результатам, и в шестой, заключительной части, сделаны выводы по проведенным экспериментам.

2. Алгоритмы выделения узлов с дополнительной функциональностью

Первоначально методы упорядочивания движения телекоммуникационного трафика в пиринговых сетях, были предложены для файлообменной сети Gnutella [5, 6] путем перенаправления данных через узлы, лежащие на быстрых каналах. Принцип работы состоял в том, что узел накапливал две группы соседей: группу с пропускной способностью меньшей, чем у него, и группу узлов с пропускной способностью большей, чем у заданного узла. От узлов с меньшей пропускной способностью данные получались и перенаправлялись к соседним узлам с большей пропускной способностью. Таким образом, узлы с наибольшей пропускной способностью можно было отчасти считать суперпирами, т.е. узлами, где дополнительная функциональность заключается в организации

дополнительной маршрутизации трафика через эти узлы. Тем не менее, ввиду неадресного поиска данных, близкого к широковещанию, эти файлообменные сети были вытеснены сетями, основанными DHT-маршрутизации.

Другим подходом, предполагающим движение трафика через наименее загруженные узлы, является алгоритм $\hat{\theta}$ -графа, основанный на введении виртуального пространства, построенного на пропускных способностях дуг сети [7]. Для построения виртуальных координат узлов сети используется алгоритм Вивальди [8]. Близость узлов в таком координатном пространстве (в смысле евклидовой метрики) пропорциональна пропускной способности каналов связи между узлами: чем ближе друг к другу узлы, тем больше пропускная способность дуг между ними. Соответственно маршрутизация трафика будет идти по узлам, расположенным вблизи друг друга во введенном координатном пространстве. В нем же можно и выделить суперпиры для дополнительной маршрутизации, пользуясь методами кластеризации точек в евклидовом пространстве. Тем не менее, схемы, основанные на таких сетевых координатах (Network Coordinates, NC), хорошо работают только в модельных ситуациях. Запуская их на реальных сетях, не удается получить точные и стабильные координаты. Расхождения, в основном, вызываются двумя причинами: существенные отклонения величины задержки сигнала между парами узлов при последовательных измерениях; постоянные изменения во времени координат узлов, которые невозможно одновременно зафиксировать.

Были предложены алгоритмы выделения суперпиров, основанные на DHT-маршрутизации в пиринговых сетях, например, алгоритм SOLE [10], в основе которого лежит протокол Chord [9]. Назначение суперпиров производится в пространстве индексов узлов таким образом, что суперпиры равномерно покрывают пространство идентификаторов. Хотя найденные суперпиры позволяют сконцентрировать на себе поисковые сообщения и контроль за своими участками подсети, они не гарантируют передачи через себя основных потоков данных.

Упомянем еще два алгоритма работы в сети с использованием суперпиров SG-1 [11] и SG-2

[12] (Superpeer Gradients). В первом алгоритме строится оверлейная топология сети с использованием множеств суперпиров, их клиентов, множества незагруженных суперпиров и произвольно связанных узлов. В основе построения указанной топологии лежат следующие принципы:

- каждый клиент присоединен только к одному суперпиру;
- суперпиры связаны посредством оверлейной сети с произвольно устанавливаемыми связями между узлами;
- число суперпиров минимально.

Узел считается незагруженным, если его пропускная способность больше числа узлов, присоединенных к нему. Главная задача этого алгоритма – передавать клиентов от суперпиров с маленькой пропускной способностью к суперпирам с большой пропускной способностью и устранять суперпиры, которые не имеют собственных клиентов.

Алгоритм SG-2 [12] является развитием идей предыдущего алгоритма, но с использованием метрики расстояний между узлами $lat(p, q)$. Расстояние вычисляется по времени задержки с помощью алгоритма Вивальди. Построенная оверлейная сеть должна обладать следующими свойствами:

- каждый клиент присоединен к одному суперпиру;
- число клиентов суперпира пропорционально пропускной способности последнего;
- расстояние (в координатном пространстве Вивальди) между клиентом и суперпиром не превышает заданного значения tot ;
- два суперпира связаны, если расстояние (в координатном пространстве Вивальди) между ними меньше $tot + \delta$;
- число суперпиров минимально.

Цель алгоритма – покрыть виртуальное пространство узлов сети минимальным числом гиперсфер вокруг суперпиров радиусом tot . Недостатком этих алгоритмов является отсутствие гарантии, что с течением времени все узлы с максимальной пропускной способностью станут суперпирами, а остальные узлы – их клиентами.

В предлагаемой работе рассматривается метод выделения узлов с дополнительной функциональностью такой, что их пропускные спо-

собности/суммы пропускных способностей инцидентных им дуг, будут максимальны.

3. Задача выделения узлов с дополнительной функциональностью

3.1. Постановка задачи

Определим математическую модель телекоммуникационной сети. В ее основе лежит связный граф $G = (V, E)$, где V – множество узлов графа, E – множество ребер графа, соединяющих узлы. Каждому ребру $e_{ij} \in E$, $\{i, j\} \in V$ графа G поставлено в соответствие неотрицательное число $c_{ij} \geq 0$ – пропускная способность ребра. Дополнительно введем следующее понятие потока f между вершинами s и t : это неотрицательная функция на ребрах графа, обладающая следующими свойствами: поток неотрицателен для любого ребра $e_{ij} \in E$, то есть $f_{ij} \geq 0$; $\sum_k f_{ki} = \sum_j f_{ij}$, $\forall i \in V, i \neq \{s, t\}$ –

поток не накапливается в промежуточных узлах между s и t ; $f_{ij} \leq c_{ij}$, $e_{ij} \in E$ – поток по ребру не превышает его пропускной способности. Остаточная пропускная способность ребра e_{ij} определяется как разность пропускной способности ребра и потока по нему, то есть $c_{ij}^f = c_{ij} - f_{ij}$. Остаточной сетью называется сеть $G^f = (V, E^f)$, получаемая из исходного графа G , в которой остаются ребра с положительной остаточной пропускной способностью. В каждый момент времени требуется проложить потоки между парами вершин $\{s_1, t_1\}, \dots, \{s_n, t_n\}$ с величинами $\{d_1, d_2, \dots, d_n\}$. Маршрутизация потоков в сети осуществляется через несколько выделенных узлов – суперпиров Sp (*SuperPeers*), и для любой пары, $\{s_i, t_i\}$ каждый простой поток между ними проходит через один и более суперпиров. Более развернуто, поток продуктов от узла абонента проходят через локальный выделенный узел (суперпир), далее по подсети суперпиров, с применением алгоритмов последовательной динамической

маршрутизации, и с окончанием на узле адресате, который также обслуживается одним из суперпиров. То есть два узла сети общаются между собой через сеть суперузлов (Рис. 1).

В данном случае, первым уровнем маршрутизации в сети будет прокладка пути от узла абонента до ближайшего суперпира и обратно. Вторым уровнем маршрутизации – прокладка канала данных в подсети суперпиров. Идея подобных многоуровневых оверлейных архитектур наиболее актуальна при построении систем пирингового телевидения [16, 17], ввиду растущей популярности P2P-сетей и потокового видео [18]. Другим аналогом подграфа суперпиров в сети может являться магистральная сеть Интернета (*Internet backbone*) – главные магистрали передачи данных между взаимосвязанными сетями и основными маршрутизаторами.

Для выделения суперпиров введем такую количественную характеристику узла, как разгруженность – мера способности узла принять, обработать и передать дальше поток. Определим критерии назначения узлов сети в качестве суперпиров:

1. величина разгруженности узла должна быть не меньше определенного порога $U(p) \geq U_{Limit}$, чтобы исключить попадание перегруженных узлов в число суперпиров и тем самым не осуществлять маршрутизацию потоков через них;

2. ребра остаточной сети, соединяющие суперпиры, в исходной сети должны образовывать связную подсеть;

3. сумма разгруженностей суперпиров должна быть максимальной;

4. оставшиеся узлы присоединяются к ближайшему суперпиру;

5. минимизировать длину пути от узла до ближайшего (по числу хопов) суперпира.

Предлагаемый метод позволяет достоверно удовлетворить требованиям 1 и 3, однако, не исключено невыполнение требования 2.

3.2. Предпосылки введения метода

Приведем рассуждения, подводящие нас к реализации метода агрегации сетевых данных. Пусть m – минимальное значение разгруженности узлов в сети, а M – максимальное значение разгруженности узлов в сети. Пусть mQ –

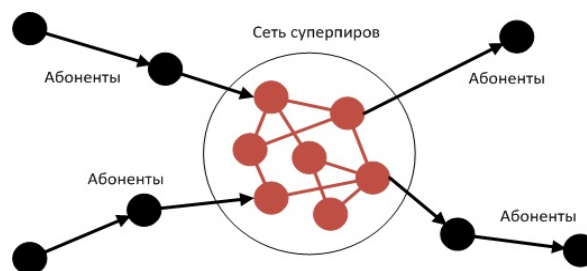


Рис. 1. Схема двухуровневой маршрутизации

пороговое значение разгруженности, и при разгруженности узла $U(p) \geq mQ$, узел p становится суперпиром.

Предположим, что удалось получить плотность распределения разгруженности узлов в сети $\rho(x)$. Тогда общее число узлов в сети будет определяться как $\int_m^M \rho(x) dx$, а число суперпиров в сети будет $\int_{mQ}^M \rho(x) dx$ и, соответственно, доля суперпиров в сети равна

$$Frac = \frac{\int_{mQ}^M \rho(x) dx}{\int_m^M \rho(x) dx}. \text{ Если функция плотности}$$

распределения может быть представлена в виде полинома $\rho(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ (степень полинома n определяется отдельно), то доля суперпиров будет равна:

$$Frac = \frac{\int_{mQ}^M \rho(x) dx}{\int_m^M \rho(x) dx} = \frac{a_0(M - mQ) + \frac{a_1}{2}(M^2 - (mQ)^2) + \dots + \frac{a_n}{n+1}(M^{n+1} - (mQ)^{n+1})}{a_0(M - m) + \frac{a_1}{2}(M^2 - m^2) + \dots + \frac{a_n}{n+1}(M^{n+1} - m^{n+1})}$$

В любом случае, долю $Frac$ суперпиров в сети можно связать с порогом вхождения mQ узла в подмножество суперпиров, при наличии $\rho(x)$. В следующем пункте будет предложен способ нахождения плотности распределения разгруженности на основе метода агрегации данных.

3.3. Метод агрегации сетевых данных

В основе работы метода агрегации данных лежит попарный обмен и усреднение данных (значений функций) между элементами связной

сети. Предположим, что дан начальный вектор чисел $w_0 = (w_{0,1} \dots w_{0,N})$. Эти значения находятся на N узлах сети и являются независимыми случайными числами с ограниченной дисперсией. Рассмотрим базовую операцию метода агрегации данных, а именно, обмен значениями функции между двумя произвольными узлами и нахождения их среднего арифметического. Для каждого узла k сети находится парный узел l с помощью операции $l = \text{getPair}(k)$, после чего производится изменение их значений по правилу: $w_k = w_l = (w_k + w_l)/2$ - операция $\text{Avg}(w_i)$. В работах [13, 14] доказывается, что при последовательном применении операций getPair и Avg , определяющих серию векторов w_1, w_2, \dots, w_n , где $w_n = \text{Avg}(w_{n-1})$, дисперсия элементов вектора w_i , равная $\sigma_i^2 = \sigma_{w_i}^2 = \frac{1}{N-1} \sum_{k=1}^N (w_{i,k} - \bar{w}_i)^2$, стремится к нулю. Это означает, что с течением времени, все узлы будут обладать одинаковыми значениями функции сети.

3.4. Применение метода агрегации данных к задаче нахождения суперпинов [15]

Для нахождения глобальных характеристик сети, таких как эмпирическая функция распределения разгруженности узлов (а также максимальное/минимальное значение разгруженности узлов в сетях, их математическое ожидание и дисперсию), необходимо соответствующим образом определить переменные, которыми будут обмениваться узлы. Пусть каждый узел сети поддерживает соответствующий вектор данных, состоящий из следующих элементов: $(TTL, n, \min U, \max U, d)$, TTL - время жизни вектора данных, n - параметр, определяющий размер сети, $\min U, \max U$ - расчет минимальной и максимальной загрузки узлов в сети, d - дополнительный вектор из b - элементов для нахождения распределения загрузки в сети (наподобие столбцовой диаграммы). В самом начале узел создает вектор: $(TTL, 1, U(p), U(p), I_p)$ (второй элемент вектора равен единице для одного узла, а для остальных его значение равно нулю), где I_p - начальная гистограмма распределения, определяемая

$$I_p(i) = \begin{cases} 0 & \text{если } U(p) < \min U + i \times \lambda \\ 1 & \text{если } U(p) \geq \min U + i \times \lambda \end{cases}$$

$0 \leq i < b$, $\lambda = \frac{\max U - \min U}{b-1}$, b - системный параметр, такой что $b > 1$. Далее для любых двух вершин p и q производится обмен и модификация

их векторов. Причем модификация векторов данных производится следующим образом:

$$\left(\frac{TTL_p + TTL_q}{2} - 1, \frac{n_p + n_q}{2}, \min(\min U_p, \min U_q), \max(\max U_p, \max U_q), \frac{d_p + d_q}{2} \right).$$

В завершение процесса агрегации каждый узел p рассчитывает системные характеристики: число узлов в сети $N = \frac{1}{n_p}$, минимальное значение разгруженности узлов в сети $\min U = \min U_p$, максимальное значение разгруженности узлов в сети $\max U = \max U_p$; вычисляется гистограмма распределения количества узлов от их величины разгруженности D посредством следующей операции:

$$D = \frac{d_p(i)}{n_p} \quad 0 \leq i < b.$$

Данная гистограмма D и является аппроксимацией плотности распределения разгруженностей узлов $\rho(x)$, указанной в разделе 3.2, при этом вычисление D производится независимо на каждом узле сети. Далее при наличии информации о необходимой доле суперпинов в сети, каждый узел может определить по гистограмме распределения количества узлов от разгруженности пороговое значение разгруженности mQ и определиться с вхождением в семейство суперпинов. Отметим, что хотя такой подход к нахождению гистограммы D и позволяет адаптивно определять порог mQ , связность подграфа получаемых суперпинов приходится определять отдельно.

3.5. Определение функции разгруженности узлов

В ходе работы было предложено четыре различных функции разгруженности узлов:

$$1. U(p) = \frac{\sum_i c_{pi} - f_{pi}}{\sum_i c_{pi}}. \quad (1)$$

Числитель является суммой остаточных пропускных способностей ребер исходящих/входящих/всех из узла, а знаменатель - сумма исходных пропускных способностей исходящих/входящих/всех ребер узла.

$$2. U(p) = \lambda_1 * \exp\left(\frac{rescap(p)}{\max_p(rescap(p))}\right) + \lambda_2 * \exp\left(\frac{d(p)}{\max_p(d(p))}\right). \quad (2)$$

$rescap(p)$ - сумма остаточных пропускных способностей ребер инцидентных вершине, $d(p)$ - степень рассматриваемой вершины, обе эти величины нормированы на максимальные значения этих величин, встречающиеся в сети; λ_1 и λ_2 - нормирующие коэффициенты, определяющие решающий вклад степени или остаточной пропускной способности вершины.

$$3. U(p) = \frac{\sum_i c_{pi} - f_{pi}}{\sum paths(p)}. \quad (3)$$

Числитель является суммой остаточных пропускных способностей ребер исходящих/входящих/всех из узла, а знаменатель - сумма потоков (выраженных через число простых путей), проходящих через данную вершину.

$$4. U(p) = \lambda_1 * \left(\frac{rescap(p)}{\max_p(rescap(p))}\right) + \lambda_2 * \left(\frac{cycle_3(p)}{\max_p(cycle_3(p))}\right) \quad (4)$$

$rescap(p)$ - сумма остаточных пропускных способностей ребер инцидентных вершине, $cycle_3(p)$ - число трехзвенных циклов, начинающихся и заканчивающихся в узле p ; λ_1 и λ_2 - нормирующие коэффициенты.

Все указанные функции разгруженности были использованы в последующих численных

экспериментах для определения наиболее подходящей для последующей реализации алгоритма двухуровневой маршрутизации.

4. Описание организации вычислительных экспериментов

В настоящем исследовании представлены результаты численных экспериментов, аналогичных таковым в работе [1], причем численному исследованию подверглись два множества сетей - сети с кластерной и стохастической топологиями.

Для каждого вида функции разгруженности (1) - (4) проводилось две серии численных экспериментов, аналогично работе [1], а именно:

1. нахождение общего числа суперпиров на каждом шаге при поступающей удовлетворенной заявке в сети, а также числа связанных подграфов, образуемых найденными суперпирами (Эксперимент 1);

2. нахождение связанного подграфа суперпиров, кластеризация узлов вокруг ближайших суперпиров и анализ степени покрытия кластеризованными вершинами всей сети. (Эксперимент 2).

Во втором эксперименте, помимо собственно бинаризации, применялся также алгоритм объединения несвязных подсетей суперпиров в связанное подмножество и производилась кластеризация сети, то есть соотнесение каждому суперпиру некоторого количества обычных узлов, которые в дальнейшем обслуживались бы им (т.е. потоки из/в обычный узел проходили бы через данный суперпир). Подробное описание этого алгоритма приведено в [1].

Условия экспериментов были подобраны таким образом, чтобы можно было сравнивать результаты выделения суперпиров методом агрегации сетевых данных (настоящая работа) и методом минимальных разрезов на графе признаков [1]. Такое сравнение также приведено в данной работе.

Опишем модели телекоммуникационных сетей, использовавшихся в экспериментах. В каждой сети были заданы пары корреспондирующих вершин (полюсов). В каждый момент времени произвольная пара корреспондирующих вершин инициирует заявку - требование проложить по сети между этими двумя полюсами единичный поток, имеющий конечное

время жизни. По истечении времени жизни заявка снимается, соответствующий поток исчезает, высвобождая ресурс сети, в данном случае остаточную пропускную способность ребер сети. Если же не удастся проложить поток между заданной парой вершин ввиду занятости ресурсов сети, то заявка считается неудовлетворенной и сбрасывается. Выполненные численные эксперименты основывались на результатах численного моделирования по заполнению сетей потоками последовательных заявок согласно описанному выше сценарию [3, 4].

В первом эксперименте, наиболее приемлемым считался результат, при котором число связанных подграфов суперпириров минимально и минимальна дисперсия числа выделяемых суперпириров и их подграфов. Отметим, что среднее число суперпириров, задаваемое в каждой сети, составляло треть от общего числа узлов в каждой сети.

Во втором эксперименте наиболее приемлемым считался результат, при котором покрытие сети кластерами узлов максимально.

Сравнение результатов выделения суперпириров двумя методами состояло в том, чтобы на каждой операции выделения суперпириров узнать число совпавших суперпириров SPC при выделении их методами минимального разреза на графе признаков и агрегации данных; а также найти отношение этого числа к общему числу суперпириров, найденных обоими методами. Последние обозначаются $Clst1$ и $Clst2$, соответственно упомянутые отношения обозначаются как $SPC/Clst1$ и $SPC/Clst2$.

Зная, сколько суперпириров могут выделить оба метода, и учитывая количество связанных подграфов суперпириров, можно определить, какой способ и какая функция загрузки дают лучший результат, т.е. меньшее число суперпириров при меньшем же числе связанных компонентов.

В проведенных экспериментах рассматривались сети с кластерной [3] и стохастической топологиями [4]. В сетях с кластерной топологией количество вершин колебалось от 30 до 45, плотность заполнения сети ребрами колебалась от 10% до 20%. В сетях со стохастической топологией количество вершин колебалось от 30

до 45, плотность заполнения сети ребрами колебалась от 10% до 20%. В данном случае под кластерной топологией сети следует понимать следующее: были даны связанные графы со сгруппированными узлами, внутри каждой группы узлы были соединены ребрами с большой пропускной способностью (4-5 кластеров узлов на сеть). Такие кластеры узлов между собой были связаны ребрами меньших пропускных способностей. Подчеркнем еще раз, что кластеризация в данной работе проводилась по такой характеристике узлов, как разгруженность, и образованные кластеры из узлов и суперпириров никак не были связаны с кластерами в топологии исходных сетей. Длина очереди заявок равнялась 45000, время жизни заявок колебалось от 10 до 1000 единиц времени с математическим ожиданием в 300 единиц времени. В каждом классе бралось по 60 сетей.

Для самого алгоритма агрегации выбирались следующие параметры: время жизни вектора данных на узле $TTL = 100$ единиц времени, мелкость разбиения b диапазона значений загрузки по сети равнялось 10.

5. Результаты проведенных численных экспериментов

Приведем результаты эксперимента 1 для различных алгоритмов последовательного заполнения сетей. На Рис.2 и Рис.3 представлены результаты по нахождению среднего числа суперпириров, выделяемых на сетях для каждой функции загрузки узла. Ввиду того, что результаты численных экспериментов слабо зависели от исходного алгоритма последовательного заполнения сети потоками продуктов, то здесь и далее графики будут приведены для субоптимального минимально-разрезного алгоритма последовательного заполнения сети. В эксперименте по нахождению общего среднего числа суперпириров, выделяемых методом агрегации данных, число суперпириров составило одну треть от общего количества узлов. Суперпириры выделялись на всех остаточных сетях, полученных в ходе последовательного заполнения сетей потоками продуктов, и таким образом, на графиках приведены средние значения вместе с дисперсией числа суперпириров.

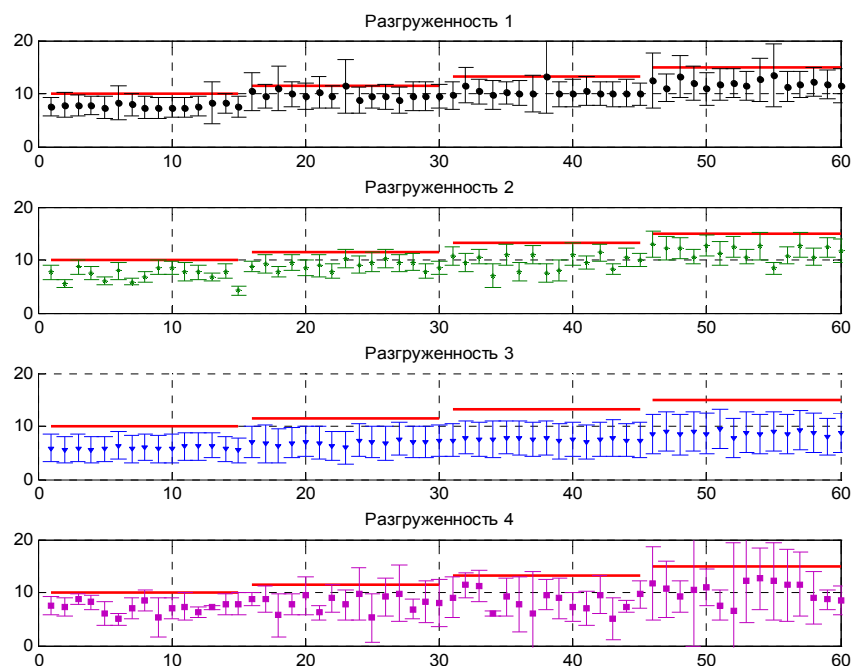


Рис. 2. Среднее число найденных суперпиков для каждой сети (кластерная топология)

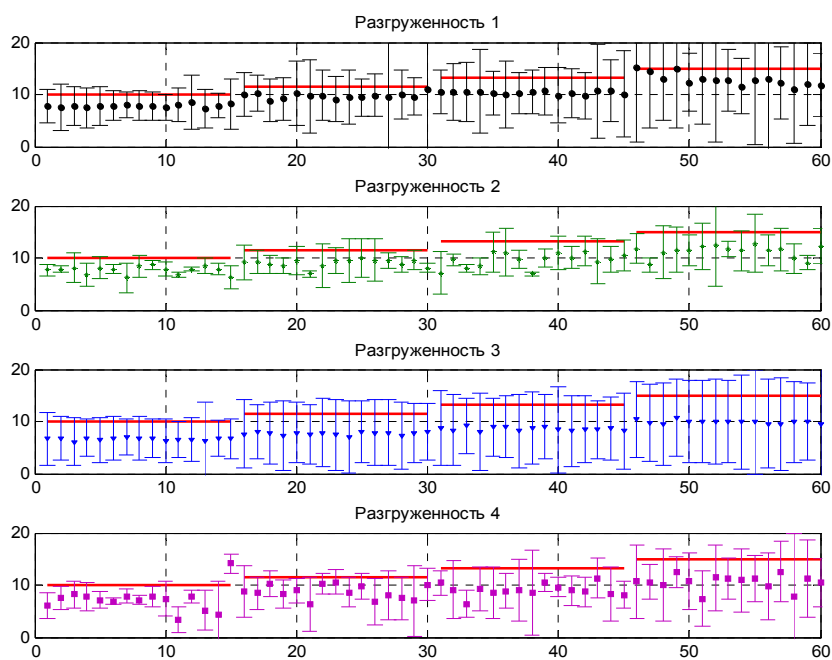


Рис. 3. Среднее число найденных суперпиков для каждой сети (стохастическая топология)

Горизонтальными линиями показано необходимое число суперпиков ($1/3$ от общего числа узлов сети). Ось абсцисс – номер сети, ось ординат – число выделяемых суперпиков.

В среднем, первая, вторая и четвертая функции разгруженности выделяют число, близкое к требуемому числу суперпиков, т.е. $1/3$ от общего числа узлов в сети, одинаково на сетях

с кластерной и стохастической топологиями. Хотя первая (1) и четвертая (4) функции разгруженности дают большие отклонения от необходимого числа суперпиров (для первой функции, сети № 46-49 стохастической топологии, сети №11-15 стохастической топологии для четвертой функции). Число суперпиров, получаемых 3-й функцией распределения, несколько меньше, что объясняется различными гистограммами распределения D (Рис. 4).

Указанный порог округления и объясняет неточность и меньшее значение числа суперпиров

в полученных результатах. В подтверждение предыдущим наблюдениям, результаты показывают большую дисперсию у четвертой функции разгруженности на сетях с кластерной топологией и у первой функции разгруженности на сетях со стохастической топологией. На Рис.5 и Рис.6 приведены результаты по среднему числу связанных подграфов суперпиров в сетях. Наиболее удовлетворительным результатом будет нахождение каждый раз такого набора суперпиров, который образует связный подграф, т.е. число связанных элементов равно 1.

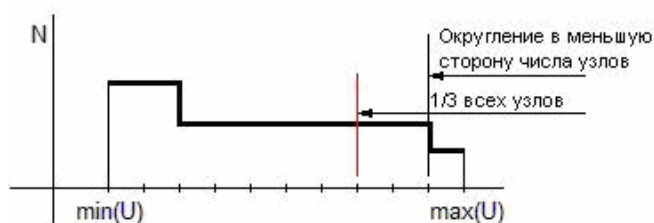


Рис. 4. Эмпирическое распределение узлов по степени разгруженности

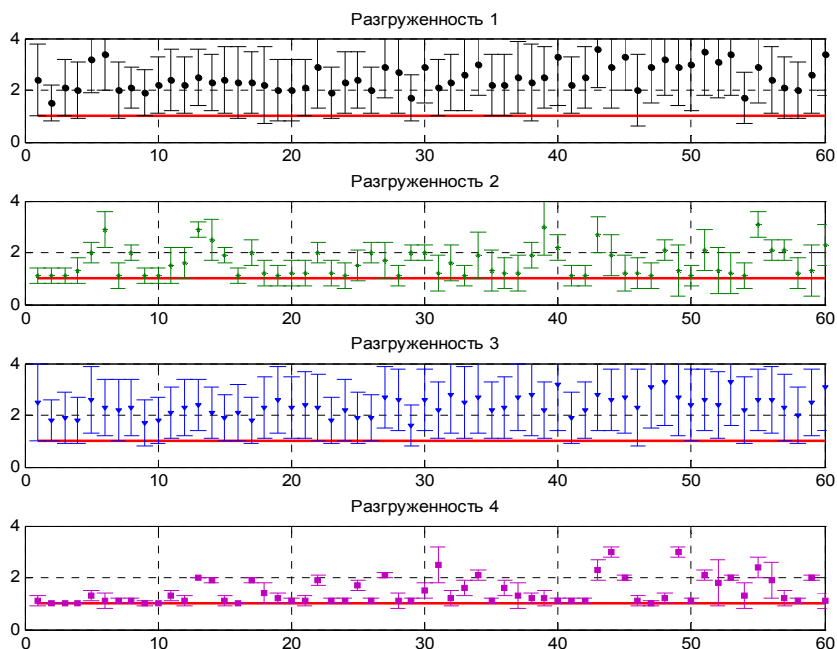


Рис. 5. Среднее число связанных подграфов суперпиров (кластерная топология)

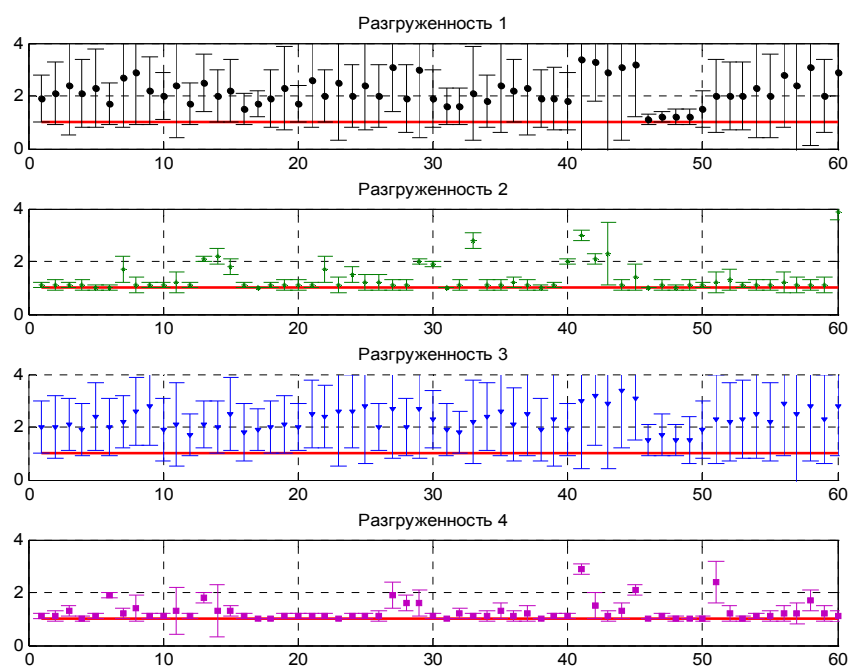


Рис. 6. Среднее число связанных подграфов суперпириов (стохастическая топология)

Результаты численных экспериментов показывают, что наименьшее число связанных подграфов суперпириов получаются при использовании функций разгруженности сети (2) и (4) для обоих классов сетей. В общем, в 63% сетей с кластерной топологией и в 70% сетей со стохастической топологией удалось получить среднее число связанных подграфов суперпириов меньше 1.5, т.е. можно предполагать, что суперпири образуют связную подсеть. Также можно заметить, что на результат слабо влияет способ образования остаточной сети, т.е. выбранный алгоритм последовательного заполнения сети потоками продуктов.

Приведем результаты эксперимента 2 для различных алгоритмов последовательного заполнения сетей. Во втором эксперименте наиболее удовлетворительным считается резуль-

тат, при котором покрытие сети кластерами узлов максимально. Для каждой функции разгруженности приведем сводные таблицы доли полностью покрытых кластерами сетей.

Из Табл. 2 видно, что на сетях со стохастической топологией доля полностью покрытых кластерами узлов сетей выше, чем на сетях с кластерной топологией (особенно это заметно в третьей функции разгруженности на кластерных сетях). Если для сетей с кластерной топологией характерная доля полностью покрытых сетей при радиусе кластера в 2 ребра равна 50%, то на стохастических сетях характерная доля покрытых сетей – 68-70%.

Тем не менее, функции разгруженности (2) и (4) лучше всего удовлетворяют требованиям к результатам экспериментов, указанных в частях 3 и 4, и обладают следующими свойствами:

Табл. 1. Сети с кластерной топологией

	Алгоритм	Простой	Дуговой	С.МинРаз	А.МинРаз	Гибрид	С.Дуговой
Разгруженность 1	Радиус 2	55%	11.7%	48.3%	50%	51.7%	31.7%
	Радиус 3	95%	90%	95%	95%	95%	95%
Разгруженность 2	Радиус 2	51.7%	53.3%	51.7%	51.7%	51.7%	51.7%
	Радиус 3	95%	95%	95%	95%	95%	95%
Разгруженность 3	Радиус 2	3.3%	1.7%	3.3%	3.3%	3.3%	3.3%
	Радиус 3	53.3%	55%	53.3%	55%	55%	56.7%
Разгруженность 4	Радиус 2	41.7%	40%	41.7%	41.7%	43.3%	41.7%
	Радиус 3	93.3%	93.3%	93.3%	93.3%	93.3%	93.3%

Табл. 2. Сети со стохастической топологией

	Алгоритм	Простой	Дуговой	С.МинРаз	А.МинРаз	Гибрид	С.Дуговой
Разгруженность 1	Радиус 2	73.3%	66.7%	73.3%	75%	73.3%	63.3%
	Радиус 3	98.3%	98.3%	98.3%	98.3%	98.3%	98.3%
Разгруженность 2	Радиус 2	71.7%	76.7%	71.7%	71.7%	71.7%	73.3%
	Радиус 3	98.3%	98.3%	98.3%	98.3%	98.3%	98.3%
Разгруженность 3	Радиус 2	53.3%	53.3%	51.7%	51.7%	55%	53.3%
	Радиус 3	91.7%	93.3%	91.7%	88.3%	95%	91.7%
Разгруженность 4	Радиус 2	68.3%	65%	68.3%	68.3%	68.3%	66.7%
	Радиус 3	95%	95%	93.3%	95%	93.3%	93.3%

1. меньшее число выделяемых связанных подграфов суперпиров по сравнению с функциями (1) и (3);

2. функция (2) обладает малой дисперсией числа выделяемых суперпиров и подграфов суперпиров;

3. функция (2) дает большую долю полностью покрытых сетей кластерами узлов при малых радиусах кластеров; функции (1) и (4) имеют примерно одинаковую долю покрытых сетей при разных радиусах кластеров узлов.

Наконец, результаты по сравнению двух методов выделения суперпиров - метода минимальных разрезов на графе признаков и метода агрегации сетевых данных – представлены на Рис.7. Результаты получены на серии сетей со стохастической топологией с количеством вершин от 30 до 45, плотностью заполнения ребрами от 10% до

20%. Представлены отношения SPC/C_{lst1} и SPC/C_{lst2} (ось ординат) для разных функций разгруженности узлов, в зависимости от номера сети (ось абсцисс).

Ранее [1] было показано, что при использовании метода минимальных разрезов на графе признаков наилучшим образом показала себя вторая функция разгруженности: несмотря на большое количество выделяемых суперпиров (в среднем >10 узлов), среднее число связанных подграфов было равно единице. В методе агрегации данных при достаточно фиксированном числе суперпиров (1/3 от общего числа узлов в сети), функции разгруженности (2) и (4) дали, в среднем, наименьшее число связанных подграфов суперпиров.

По результатам экспериментов обнаружено, что функция разгруженности (4) для метода

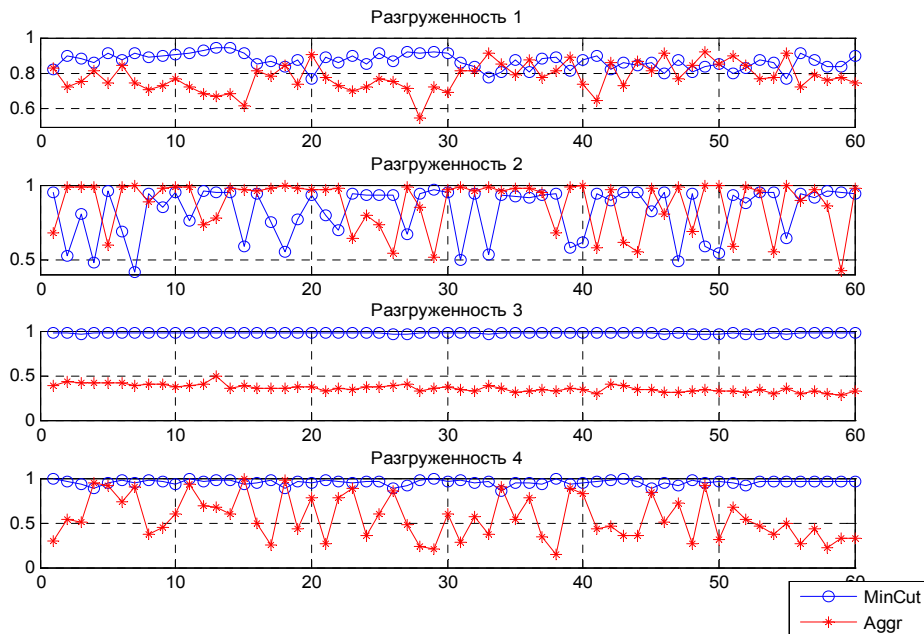


Рис. 7. Доля совпавших суперпиров, относительно общего количества выделенных суперпиров в двух алгоритмах выделения суперпиров

минимального разреза дает меньшее число суперпиров, при том что число связных подграфов суперпиров, в большинстве случаев, равно одному. Для функции (2) наблюдается, что в 50% случаев метод агрегации данных выделяет меньше суперпиров, чем метод минимальных разрезов, хотя использование этой функции дает наименьшее число связных подграфов суперпиров. Использование функции (3) показывает, что метод агрегации данных дает избыточные узлы с дополнительной функциональностью по сравнению с методом минимальных разрезов на графе признаков. Первая функция позволяет понять, что оба метода выделяют некоторое множество избыточных суперпиров, при том что количество связных подграфов принимает значения от 2 до 4-х.

Таким образом, функции разгруженности, построенные с учетом геометрических особенностей графа сети, можно рекомендовать при построении алгоритмов двухуровневой маршрутизации в сетях передачи данных. На использованных массивах сетей не удалось получить очевидных недостатков метода агрегации сетевых данных перед методом минимального разреза на графе признаков. Хотя априорные недостатки метода агрегации были указаны выше.

Заключение

В данной работе была рассмотрена задача выделения связной подсети узлов с дополнительной функциональностью в сетях передачи данных и присоединения оставшихся узлов к ним. Были предложены четыре функции разгруженности узлов, по которым производилось нахождение кандидатов в узлы с дополнительной функциональностью. Был предложен алгоритм их выделения, основанный на методе агрегации сетевых данных.

Проведенные численные эксперименты имели целью:

- найти общее число суперпиров, оценить их среднее число и дисперсию для каждой сети;
- найти количество связных подграфов, обслуживаемых суперпирами;
- при выделении связного подграфа суперпиров и определении множества оставшихся

узлов, обслуживаемых каждым из суперпиров, найти степень покрытия сети образовавшимися группами узлов.

По результатам проведенных экспериментов для функции разгруженности (2) дисперсия числа суперпиров была минимальна. В 63% сетей с кластерной топологией и 70% сетей со стохастической топологией с помощью функций разгруженности (2) и (4) удалось получить связные подграфы суперпиров. Функции разгруженности (1), (2) и (4) выделяют требуемое число суперпиров, т.е. 1/3 от общего числа узлов в сети, одинаково на сетях с обоими типами топологий. Результаты в Табл. 1 и Табл. 2 показывают, что на сетях со стохастической топологией доля полностью покрытых сетей кластерами узлов выше, чем на сетях с кластерной топологией.

Несмотря на то, что метод агрегации уступает методу минимальных разрезов на графе признаков, так как он не гарантирует связность подграфа суперпиров, он обладает такими преимуществами, как:

1. дает возможность задать точное число суперпиров в сети;
2. требует нетривиально настраиваемых параметров для работы алгоритма;
3. является полностью распределенным алгоритмом;
4. данные не требуется пересылать в один регулирующий узел для принятия решения о назначении узлов суперпирами.

Число найденных с помощью метода агрегации данных суперпиров не зависит от плотности заполнения дугами графа сети, несмотря на то, что две метрики разгруженности вычисляются с использованием такого показателя, как степень вершины.

Литература

1. Гринберг Я.Р., Курочкин И.И., Корх А.В. Алгоритм кластеризации сетей передачи данных / Информационные технологии и вычислительные системы, - Т.3 - М.: УРСС, 2012, с.18-30.
2. Гринберг Я.Р., Курочкин И.И. Математическое моделирование динамического последовательного заполнения сетей потоками связи / Проблемы вычислений в распределенной среде: Труды ИСА РАН / Под ред. С.В. Емельянова, А.П. Афанасьева – Т.46 - М.: КРАСАНД, 2009, с.233-258.

3. Гринберг Я.Р., Курочкин И.И. Исследование результатов математического моделирования последовательного заполнения сетей с кластерной топологией / Проблемы вычислений в распределенной среде: Труды ИСА РАН / Под ред. С.В. Емельянова, А.П. Афанасьева – Т.46 - М.: КРАСАНД, 2009, с.198-232.
4. Гринберг Я.Р., Курочкин И.И. Анализ результатов численного эксперимента по последовательному заполнению сетей со стохастической топологией / Проблемы вычислений в распределенной среде: распределенные приложения, коммуникационные системы, математические модели и оптимизация: Сборник трудов ИСА РАН / Под ред. А.П. Афанасьева – Т.25 - М.: КомКнига, 2006, с.99-128.
5. Eng Keong Lua, Jon Crowcroft, and Marcelo Pias. A survey and comparison of peer-to-peer overlay network schemes / Univ. of Cambridge, Ravi Sharma, Nanyang Technological University, Steven Lim, Microsoft Asia. IEEE Communications, 2005.
6. Yatin Chawathe, Sylvia Ratnasamy, Lee Breslau, Nick Lanham, Scott Shenker. Making Gnutella like P2P Systems Scalable / SIGCOMM'03, August 25–29, 2003, Karlsruhe, Germany.
7. J. Ledlie, M. Mitzenmacher, M. Seltzer. Wired Geometric Routing / In Proceedings of IPTPS 2007.
8. Frank Dabek, Russ Cox, Frans Kaashoek, Robert Morris. Vivaldi: A decentralized Network Coordinate System. / SIGCOMM'04, Aug. 30-Sept. 3, 2004, Portland, Oregon, USA.
9. Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. / MIT Laboratory for Computer Science, 2001.
10. V. Lo, D. Zhou, Y. Liu, C. Gauthier Dickey, and J. Li. Scalable supernode selection in peer-to-peer overlay networks. / In Proceedings of the Second International Workshop on Hot Topics in Peer-to-Peer Systems, p. 18-27, 2005.
11. A. Montresor. A robust protocol for building superpeer overlay topologies. / In Proceedings of the 4th International Conference on Peer-to-Peer Computing, p. 202-209. IEEE Computer Society, August 2004.
12. G. P. Jesi, A. Montresor, Ö. Babaoglu. Proximity-aware superpeer overlay topologies. / In Proceedings of the 2nd IEEE International Workshop on Self-Managed Networks, Systems, and Services, volume 3996 of Lecture Notes in Computer Science, p. 43-57. Springer, June 2006.
13. M. Jelasity and A. Montresor. Epidemic-style proactive aggregation in large overlay networks / In Proceedings of the 24th International Conference on Distributed Computing Systems, p.102-109. IEEE Computer Society, 2004.
14. M. Jelasity, A. Montresor, and O. Babaoglu. Gossip-based aggregation in large dynamic networks / ACM Transactions on Computer Systems, vol. 23, p. 219-252, August 2005.
15. J. Sacha, J. Dowling, R. Cunningham, R. Meier. Using aggregation for adaptive superpeer discovery on the gradient topology / In IEEE International Workshop on Self-Managed Networks, Systems and Services, 2006.
16. U. Abbasi, T. Ahmed. SWOR: an Architecture for P2P Scalable Video Streaming using Small World Overlay / in proceedings of IEEE CCNC, Las Vegas, 2010
17. Hao Hu, Yang Guo, Yong Liu. Mesh-based peer-to-peer layered video streaming with taxation / Proceedings of the 20th International workshop on Network and Operating Systems Support for Digital Audio & Video (NOSSDAV 2010), 2010.
18. H. Schulze, K. Mochalski. IPOQUE Internet Study 2008/2009. Research was performed by ipoque GmbH / www.ipoque.com/resources/internet-studies, 2010.

Гринберг Яков Рувимович. Центр Грид-технологий и распределенных вычислений. Кандидат технических наук. Автор 48 печатных работ. Область научных интересов: математическое моделирование потоков в сетях, алгоритмы канальной маршрутизации, задачи многопродуктовой маршрутизации. E-mail greenjak@isa.ru

Курочкин Илья Ильич. Центр Грид-технологий и распределенных вычислений. Кандидат технических наук. Автор 35 научных работ. Область научных интересов: имитационное моделирование телекоммуникационных сетей, последовательные алгоритмы на графах, динамическая канальная маршрутизация, выделение объектов на двумерных изображениях, анализ текстур на двумерных изображениях. E-mail qurochkin@gmail.com

Корх Артем Владимирович. Центр Грид-технологий и распределенных вычислений, Московский физико-технический институт. Аспирант. Автор 14 печатных работ. Область научных интересов: моделирование телекоммуникационных сетей, сегментация растровых и спектральных изображений, алгоритмы сжатия видео потоков. E-mail artkorkh@gmail.com

Алыгулиев Рамиз Магамед оглы. Институт информационных технологий НАН Азербайджана. Заведующий отделом. Доктор технических наук. Автор 90 печатных работ. Область научных интересов: распределенные вычисления, нейронные сети, эволюционные алгоритмы, оптимизация на основе роевого интеллекта. E-mail: a.ramiz@science.az

Гашимов Мамед Ариф оглы. Институт информационных технологий НАН Азербайджана. Аспирант. Автор 2 печатных работ. Область научных интересов: синтез распределенной вычислительной среды; ГРИД и облачные вычисления. E-mail: mamedhashimov@gmail.com