

Алгоритм кластеризации элементов сетей передачи данных¹

Я.Р. Гринберг, И.И. Курочкин, А.В. Корх

Аннотация. В данной работе решается задача нахождения в телекоммуникационных сетях узлов с дополнительной функциональностью, для организации двухуровневой маршрутизации. Для нахождения таких узлов решается задача введения характеристик, по которым множество узлов телекоммуникационной сети можно разделить на множество узлов с дополнительной функциональностью и на оставшиеся узлы. Рассматривается задача присоединения оставшихся узлов к узлам с дополнительной функциональностью, то есть кластеризация узлов.

Ключевые слова: сеть, маршрутизация, суперпир, бинаризация, функция разгруженности.

Введение

В данной работе рассматривается задача выделения узлов телекоммуникационных сетей для организации двухуровневой маршрутизации. Под двухуровневой маршрутизацией понимается следующее: в телекоммуникационной сети выделяется некоторое множество узлов, которым назначается дополнительная функциональность по маршрутизации трафика (далее они будут называться суперузлы или суперпиры). Маршрутизация потоков продуктов идет от узла абонента через локальный выделенный узел, далее через подсеть суперпиров и оканчивается на узле адресата, который также принадлежит одному из суперпиров. Иными словами два узла сети общаются между собой через сеть суперузлов. Задача двухуровневой маршрутизации возникает в рамках решения задачи последовательного заполнения сети потоками продуктов [14-16]. Напомним, что в указанных работах рассматривалась задача связи абонентов сети, то есть прокладки канала между полюсами, при последовательном поступлении заявок на соединения произвольных пар абонентов. В качестве решения были

предложены алгоритмы последовательного заполнения сети, основанные на различных эвристиках, а именно:

- Учет вхождения ребер сети в минимальные разрезы между различными парами полюсов (корреспондирующие абоненты). После вычисления величин минимальных разрезов с помощью различных функционалов вычислялись веса ребер, принадлежащих этим минимальным разрезам, и далее потоки прокладывались по путям минимальных стоимостей.
- Веса ребер назначаются согласно пропускной способности ребер. Далее поток прокладывается по пути минимальной стоимости.
- Всем ребрам, независимо от их пропускных способностей, назначаются единичные веса. Путь минимальной стоимости был эквивалентен кратчайшему пути.

Заметим, что маршрутизация потоков продуктов этими алгоритмами была одноуровневой, то есть путь продукта сразу вычислялся от источника к стоку без прохождения через заранее заданные узлы. Однако скорость работы заявленных алгоритмов последовательного заполнения сетей продуктами сильно замедлялась

¹ Поддержка 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), программа фундаментальных исследований РАН № 14.

на сетях больших размеров ($>10^4$ узлов) при последовательном обслуживании каждой поступающей заявки, особенно при нахождении всех ребер, входящих в минимальные разрезы между различными парами полюсов. Для преодоления возникающих вычислительных сложностей было предложено использовать двухуровневую маршрутизацию при обслуживании поступающих заявок. Подобную схему маршрутизации можно наблюдать в пиринговых сетях (сети в основе построения и функционирования которых лежит принцип равноправия участников, одним из признаков которого является совмещение узлом сети ролей и клиента и сервера в отличие от чистой клиент-серверной модели [1]). Существует несколько типов организации пиринговых сетей, а именно:

- построение децентрализованной сети - нет выделенных узлов с дополнительной функциональностью;
- построение смешанной (гибридной) сети - есть некоторые узлы с дополнительной функциональностью;
- построение централизованной сети - использование центрального сервера.

В данном случае нахождение суперпиров и обычных узлов используется в пиринговых сетях со смешанной топологией; примерами таких сетей являются Fidonet, FastTrack, JXTA-сети, Skype, DHT-сети, Kad Network [1-3].

В данной работе решается подзадача задачи двухуровневой маршрутизации, а именно определение критерия выделения суперпиров, нахождения их в сети и присоединения к ним оставшихся узлов сети, то есть нахождения кластеров узлов сети вокруг выделенных суперпиров.

Дальнейшая структура работы будет следующей: во второй части будет дан краткий обзор существующих на данный день алгоритмов выделения узлов с дополнительной функциональностью и организации двухуровневой маршрутизации. В третьей части будет дана постановка задачи, определены особенности решаемой нами задачи и дано описание алгоритма. В четвертой части читателю будет представлено математическое описание экспериментов, затем их результаты и в шестой, заключительной части, сделаны выводы по проведенным экспериментам.

1. Существующие алгоритмы выделения суперпиров и организации двухуровневой маршрутизации

На сегодняшний день существует несколько методов нахождения суперпиров в сети, построения вокруг них узлов - пользователей и связывания суперпиров в единую оверлейную сеть. В частности, одним из таких алгоритмов является алгоритм $\hat{\theta}$ -графа, основанного на введении виртуального пространства, построенного на пропускных способностях дуг сети [4]. Для построения виртуальных координат узлов сети используется алгоритм Вивальди [5]. В основе его работы лежит модель, рассматривающая сеть как набор точек, связанных между собой пружинами, если в реальной сети узлы были напрямую соединены. Каждый узел сети хранит данные о своих координатах \bar{x}_i и «уверенности» в своей координате $w_i \in (0,1)$. На каждой итерации алгоритма узлы меняют свое положение в виртуальном пространстве координат в зависимости от сжатия или растяжения "пружины" между ними. Величина сжатия/растяжения пружин задается выражением $|\|\bar{x}_i - \bar{x}_j\| - l_{ij}|$, где l_{ij} - измерение времени ожидания между двумя узлами i и j . Коэффициент жесткости "пружин" вычисляется в зависимости от параметра w_i и смещения текущих координат узла от его оптимального положения. Что касается самого алгоритма $\hat{\theta}$ -графа, то маршрутизация осуществляется по наиболее близкорасположенным во введенном координатном пространстве узлам. Степень близости определяется по полученным координатам узлов сети в евклидовой метрике. Соответственно, в таком пространстве можно производить кластеризацию близкорасположенных узлов. Однако схемы, основанные на сетевых координатах (Network Coordinates, NC), хорошо работают только в модели. Будучи запущенными на реальных сетях, не удается получить точные и стабильные координаты. Расхождения, в основном, вызываются двумя причинами: 1) отклонением более чем на порядок, при измерении задержек между парами узлов; 2) свойственной задержкам неспособности точно удовлетворять

неравенству треугольника, что вызывает осцилляции.

Одним из методов кластеризации точек в пространстве, используемых в реальных алгоритмах маршрутизации, является применение кривых Гильберта [6, 7]. Данный метод нашел свое применение при реализации протокола eCAN [8]. Кривая Гильберта представляет собой пример пространственно-заполняющей кривой (space-filling curve) в k -мерном квадрате со стороной, равной степени двойки. Эта кривая проходит через все точки сетки по одному разу без каких-либо самопересечений. Кривая имеет два свободных конца, которые могут быть соединены с концами других кривых. Базовая кривая называется кривой первого уровня. Для получения кривой уровня $i+1$ каждая вершина кривой i заменяется на кривую первого порядка, которая может быть повернута соответствующим образом (Рис. 1).

Соответственно, в кластеры объединяются элементы, лежащие в соседних ячейках по ходу движения по кривой Гильберта. В качестве центральных элементов кластеров выбираются те элементы, которые получаются усреднением свойств всех входящих в данный кластер представителей. Координатное метрическое евклидово пространство строится по аналогии с Плакстон-алгоритмом (Plaxton algorithm) [9], где каждому узлу присвоена n -битная метка. Эта метка может быть разделена на l частей, в каждой из которой $w = n/l$ бит. Координаты назначаются в зависимости от значений w . Сама метка n вычисляется как хэш-функция от сетевого адреса узла.

Наконец, представим еще один алгоритм выделения, основанный на DHT-маршрутизации в пиринговых сетях. Более точно, в основе алгоритма лежит протокол Chord [10]. Основная идея алгоритма SOLE [11] заключается в назначении K идентификаторов для суперпиров, а узлы с наиболее близкими идентификаторами становятся суперпирами. Номера идентификаторов для суперпиров задаются по следующему правилу:

$$0, \left\lfloor \frac{2^m}{K} \right\rfloor, \left\lfloor 2 \frac{2^m}{K} \right\rfloor, \left\lfloor 3 \frac{2^m}{K} \right\rfloor, \dots, \left\lfloor (K-1) \frac{2^m}{K} \right\rfloor,$$

где 2^m - размер пространства идентификаторов.

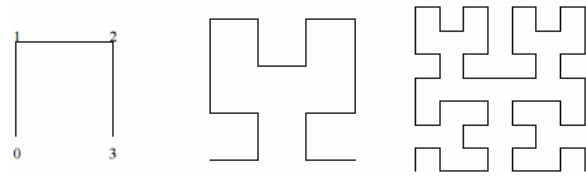


Рис. 1 Кривые Гильберта 1-го, 2-го и 3-го уровней соответственно

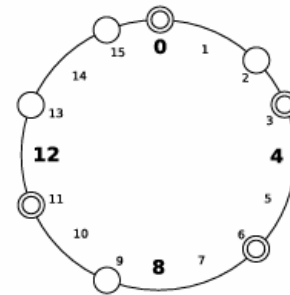


Рис. 2. Распределение суперпиров для идентификаторов 0, 4, 8, 12

На Рис.2 показано распределение суперпиров для идентификаторов 0, 4, 8, 12. В качестве суперпиров были выбраны соответственно вершины 0, 3, 6 и 11.

Данное распределение идентификаторов суперпиров разбивает все пространство идентификаторов на K равных частей, а расстояние от узла до ближайшего суперпира равно $O(\log(N/K))$, N - число узлов в сети.

Недостатком двух последних методов нахождения суперпиров и узлов вокруг них является то, что никак не учитываются такие характеристики, как загруженность узлов, пропускные способности каналов связи между узлами в сети, а назначенные для локальной маршрутизации трафика узлы могут не справиться с передачей данных. Предлагаемый в данной работе алгоритм выделения суперпиров, наоборот, учитывает требования по пропускным способностям элементов в сети.

2. Выделение суперпиров в сети. Постановка задачи, определение критерия

2.1. Постановка задачи

В самом начале 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 понятие потока F между вершинами s и t , неотрицательную функцию на ребрах графа, обладающую следующими свойствами: поток неотрицателен для любого ребра $e_{ij} \in E$, то есть $f_{ij} \geq 0$; $\sum_k f_{ki} = \sum_j f_{ij}$, $\forall i \in V, i \neq \{s, t\}$,

следовательно 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 ней, 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.

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

1. Величина загруженности узла должна быть не меньше определенного порога $U(p) \geq U_{limit}$, чтобы исключить попадание загруженных узлов в число суперпиров и тем самым осуществлять маршрутизацию потоков через них.

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

3. Сумма загруженностей суперпиров должна быть максимальной $\max \sum_{i \in Sp} U_i(p)$

4. Оставшиеся узлы присоединяются к ближайшему суперпиру.

5. Длина пути от узла до ближайшего (по числу хопов) суперпира $\min l(v, Sp)$ должна быть минимизирована.

Предлагаемый подход выделения суперпиров предполагает выполнение условий 1-2, а также частично условия 3. В качестве алгоритма пороговой разделения объектов множества, в нашем случае узлов сети, на два класса, было предложено использовать метод нахождения минимальных разрезов на графе признаков элементов множества [12]. В частности, структура графа признаков совпадает со структурой, в которую связаны элементы, то есть со структурой сети. Пропускные способности ребер назначаются исходя из численных значений характеристик элементов множества.

Последующее присоединение оставшихся узлов к найденным суперпирам и, таким образом, создание кластеров узлов в сети осуществляется путем присоединения к каждому суперпиру ближайших узлов на расстоянии одной, двух и более дуг.

2.2. Алгоритм разделения узлов сети на два класса

Для решения подобной задачи разделения множества объектов на два класса (в дальнейшем будем называть подобное деление бинаризацией) предполагается использовать метод минимизации энергии через минимальные разрезы на графе [12, 13]. Остановимся на подробном описании этого метода.

Предполагается, что элементы множества связаны между собой и при признаковой бинаризации необходимо учитывать соседство элементов. Каждому элементу (в нашем случае узлу сети) $p \in P$ приписывается некоторая пометка из конечного множества L . Общая задача звучит так: найти такую расстановку пометок f для каждого узла $p \in P$ из множества $f_p \in L$, что f может быть представлена некоторой функцией, соответствующей наблюдаемым данным.

Тогда задача признаковой бинаризации может быть сформулирована в терминах минимизации энергетической функции:

$$E(f) = E_{smooth}(f) + E_{data}(f) \quad (2.2.1)$$

где $E_{smooth}(f)$ определяет негладкость функции, а слагаемое $E_{data}(f)$ отличие расстановки пометок от правильного разбиения. Без ограничения общности будем считать, что $E(f) = \sum_{\{p,q\} \in N} V_{p,q}(f_p, f_q) + \sum_{p \in P} D_p(f_p)$, т.е. $E_{smooth}(f)$ определяется соседними элементами (узлами, соединенными ребром), а $E_{data}(f)$ означает непосредственно сравнение разгруженности узла и значений разгруженности, присущей типичным супернодам и обычным узлам. Также имеют место следующие соотношения

$$V(\alpha, \beta) = 0 \Leftrightarrow \alpha = \beta$$

$$V(\alpha, \beta) = V(\beta, \alpha) \geq 0$$

$$V(\alpha, \beta) \leq V(\alpha, \gamma) + V(\gamma, \beta)$$

И основной сложностью этой задачи является вычисление глобального минимума функции, рассмотренной в 2.1. Для решения данной задачи используется алгоритм нахождения минимального разреза на графе признаков. Построим граф $G_{\alpha\beta}$, (Рис. 3), состоящий из графа исходный элементов, т.е. сети узлов, а также двух дополнительных вершин α и β , являющимися полюсами графа $G_{\alpha\beta}$, и связанных дугами со всеми остальными узлами.

Тогда очевидно, что любой разрез C на графе $G_{\alpha\beta}$ содержит ровно одно t -ребро для любого пикселя $p \in P_{\alpha\beta}$. Это естественным образом определяет следующую расстановку пометок:

$$f_p^C = \begin{cases} \alpha, & \text{если } t_p^\alpha \in C, \text{ для } p \in P_{\alpha\beta} \\ \beta, & \text{если } t_p^\beta \in C, \text{ для } p \in P_{\alpha\beta} \\ f_p, & \text{для } p \in P, p \notin P_{\alpha\beta} \end{cases} \quad (2.2.2)$$

В работе [12] доказываются следующие утверждения:

Лемма 3.2.1. Для любого разреза C и любого ребра $e_{\{p,q\}}$ справедливо следующее:

- а) Если $t_p^\alpha, t_q^\alpha \in C$ то $e_{\{p,q\}} \notin C$
- б) Если $t_p^\beta, t_q^\beta \in C$ то $e_{\{p,q\}} \notin C$
- в) Если $t_p^\beta, t_q^\alpha \in C$ то $e_{\{p,q\}} \in C$
- г) Если $t_p^\alpha, t_q^\beta \in C$ то $e_{\{p,q\}} \in C$

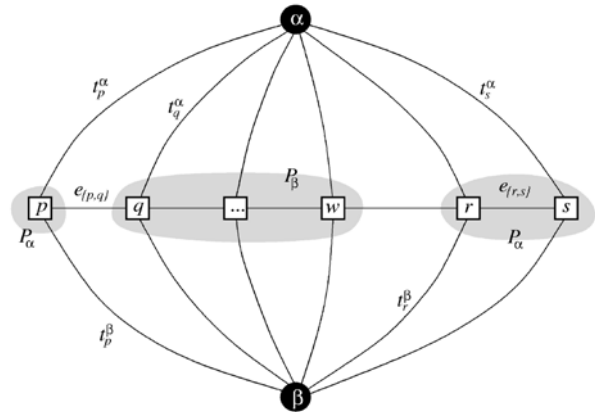


Рис 3. Граф $G_{\alpha\beta}$

Лемма 3.2.2. Для любого разреза C и любого ребра $e_{\{p,q\}}$ справедливо следующее:

$$|C \cap e_{\{p,q\}}| = V(f_p^C, f_q^C)$$

Теорема 3.2.3. Расстановка пометок, минимизирующая энергетическую функцию получается из $f = f^C$, где C минимальный разрез на графе $G_{\alpha\beta}$.

Исходя из сказанного выше, для нахождения границы требуется минимизировать функцию $E(f) = \epsilon * V(f) + D(f)$, где f - вектор меток узлов $f = (f_1, f_2, \dots, f_n)$, слагаемое $D(f)$ - это $D(f) = \sum_{p \in P} D_p(f_p)$, т.е. сумма индивидуальных штрафов, за принадлежность узла подмножеству суперузлов или обычных узлов, то есть насколько он отличается от суперузла или обычного узла по показателю разгруженности; слагаемое $V(f) = \sum_{\{p,q\} \in Neighbours} V_{\{p,q\}} * \delta(f_p, f_q)$,

где $\delta(f_p, f_q) = \begin{cases} 1 & \text{if } f_p \neq f_q \\ 0 & \text{otherwise} \end{cases}$ - дает сумму штрафов за резкость в изменении значения разгруженности соседних узлов, а ϵ - коэффициент, показывающий, насколько превалирует критерий расстановки меток по отличию от контрольных значений узла/суперузла от критерия определения границы раздела между узлами с разными метками по резкости изменения величины разгруженности соседей. То есть два различных критерия, по которым можно произвести бинаризацию, определяют два типа ребер в графе. Также данный метод бинаризации

позволяет надеяться, что выделенные в сети множества суперузлов и обычных узлов будут составлять связные подграфы на ребрах исходной сети, т.к. разрез проходит по исходной телекоммуникационной сети.

Далее стоит вопрос с определением пропускных способностей дуг образованного графа $G_{\alpha\beta}$. Поскольку в исходных материалах по бинаризации с помощью минимального разреза не было дано непосредственных указаний к тому, как определяются функции $D(f)$, $V(f)$ и коэффициент ε , то в экспериментах были предложены следующие виды:

$$D_p(f) = \exp\left(-1 * \frac{|\max Util - U(p)|}{MaxUtil - MinUtil}\right)$$

для ребер, связывающих узлы сети с полюсом, соответствующим загрузке суперпиров

$$D_p(f) = \exp\left(-1 * \frac{|\min Util - U(p)|}{MaxUtil - MinUtil}\right)$$

для ребер, связывающих узлы сети с полюсом, соответствующим загрузке рядовых узлов

$$V_{pq}(f) = \exp\left(-1 * \frac{|U(p) - U(q)|}{MaxUtil - MinUtil}\right)$$

для ребер между узлами сети.

2.3. Определение функции разгруженности узлов

Единственным исходным требованием к функции разгруженности узлов $U(p)$ является ее неотрицательность. Далее было предложено три различных функции распределения:

$$1. \quad U(p) = \frac{\sum_i c_{pi} - f_{pi}}{\sum_i c_{pi}} \quad (2.3.1).$$

Числитель представляет собой сумму остаточных пропускных способностей дуг исходящих/входящих/всех из узла, а знаменатель сумму исходных пропускных способностей исходящих/входящих/всех дуг узла.

$$2. \quad U(p) = \exp\left(\lambda_1 * \frac{rescap(p)}{\max_p(rescap(p))}\right) + \exp\left(\lambda_2 * \frac{d(p)}{\max_p(d(p))}\right) \quad (2.3.2).$$

$rescap(p)$ - сумма остаточных пропускных способностей дуг 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, определяют решающий вклад степени или остаточной пропускной способности вершины.

$$3. \quad U(p) = \frac{\sum_i c_{pi} - f_{pi}}{\sum paths(p)} \quad (2.3.3).$$

Числитель представляет собой сумму остаточных пропускных способностей дуг исходящих/входящих/всех из узла, а знаменатель сумму потоков (выраженных через простые пути), проходящих через данную вершину.

Все указанные функции разгруженности были использованы в последующих численных экспериментах для определения наиболее подходящей для последующей реализации алгоритма двухуровневой маршрутизации.

3. Организация вычислительных экспериментов и описание их параметров

Для каждого вида функции разгруженности проводилось две серии численных экспериментов:

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

2. нахождение связного 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.

Соответствующий алгоритм состоял в следующем:

1. перед началом алгоритма находится список кратчайших путей L между всеми верши-

нами графа (с помощью алгоритма Флойда-Уоршелла);

2. с помощью списка L находятся такие пары узлов между связными частями суперпиров, что их можно соединить по кратчайшим путям между собой;

3. включаются промежуточные узлы V_h , лежащие на кратчайших путях, соединяющих связные графы суперпиров, в множество суперпиров Sp , но без права образовывать вокруг себя кластеры пользовательских узлов, так как предполагается, что загрузка V_h выше, чем у "настоящих" суперпиров, и их можно использовать только для передачи трафика между суперпирами.

Наконец, последним шагом является процесс сбора узлов-пользователей вокруг выбранных суперпиров. В данном случае можно предложить следующий способ построения кластеров узлов:

```

for каждый суперпир  $s \in Sp$ 
  for  $i = 0 : \max hops$ , где  $\max hops$  -
  максимальная удаленность (в хопах) узлов
    if  $p \notin SpClust$  - узел не занят
       $p$  in  $SpClust(p)$ 
    end;
  end;
end;

```

В первом эксперименте наиболее удовлетворительным считался результат, при котором число связных подграфов суперпиров минимально, а также минимальна дисперсия числа выделяемых суперпиров и их подграфов.

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

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 единичный поток между этими двумя полюсами, 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 удается проложить поток между заданной парой

вершин, ввиду занятости ресурсов сети, то заявка считается неудовлетворенной и сбрасывается. В качестве моделей сетей и материала для экспериментов использовались модели сетей описанные в работах [14, 15].

В проведенных экспериментах рассматривались сети с кластерной топологией [16] (количество вершин от 30 до 45, плотность заполнения сети ребрами составляет 10%). В данном случае под кластерной топологией сети понимается следующее: были даны связные 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 суперпиров 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.

Работа функции кластеризации сети вокруг суперпиров представлена на Рис.4.

4. Результаты численных экспериментов

Приведем результаты эксперимента 1 для различных алгоритмов последовательного заполнения сетей.

На Рис. 5 представлены результаты по нахождению среднего числа суперпиров выделяемых на сетях для шести алгоритмов последовательного заполнения сети потоками продуктов для каждой функции разгруженности узла. По оси абсцисс стоит номер сети с кластерной топологией, по оси ординат – число суперпиров. Значком "o" отмечены точки, соответствующие функции разгруженности узлов (2.3.1), значком "*" – функции разгруженности (2.3.2), значком "∇" – функции разгруженности (2.3.3). Графики (сверху вниз) соответствующим алгоритмам последовательного заполнения сетей:

1. простой;
2. дуговой;

3. субоптимальный минимально разрезный;
4. аддитивно минимально разрезный;
5. гибридный;
6. субоптимальный дуговой.

Вторая функция разгруженности зачастую выделяет большее число суперпиров, чем при функции (2.3.1). Первая функция разгруженности, определяемая по (2.3.1) в среднем дает 7-10



Рис. 4. Схема работы алгоритма выделения суперпиров и кластеризации узлов сети

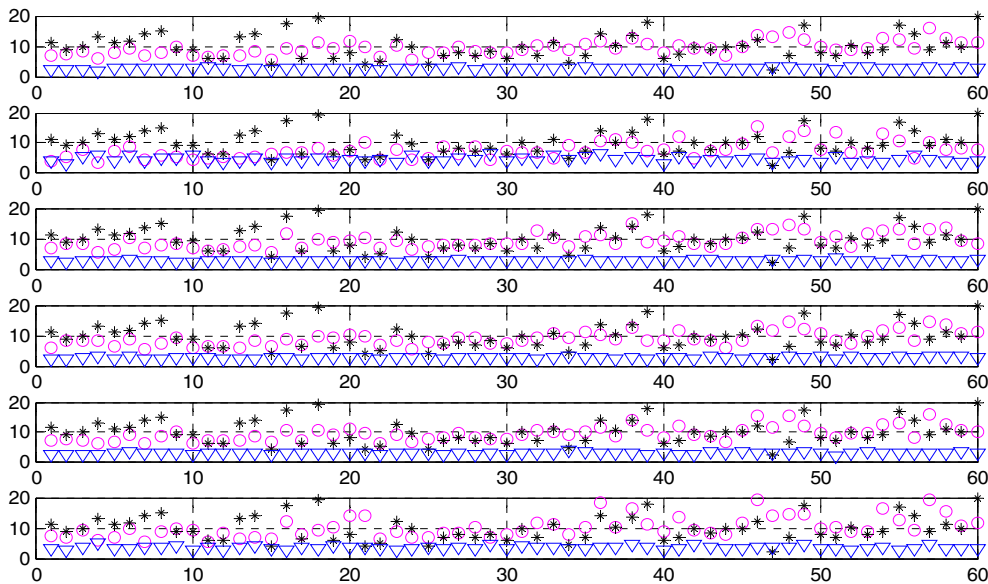


Рис. 5. Среднее число найденных суперпиров для каждой сети

суперпиров на сетях из 30 - 45 узлов. Третья функция разгруженности узлов в среднем дает 3 – 4 суперпира на сеть из 30 – 45 узлов. На Рис.6 представлены дисперсии σ^2 среднего числа значений выделяемых суперпиров.

На Рис. 6 показано, что при кластеризации с использованием функции разгруженности (2.3.1), полученный результат неустойчив (наибольший квадрат дисперсии числа суперпиров σ^2 равен 108.5 при 9.8, в среднем, суперпиров в 38 сети, дуговой алгоритм). При этом для функций разгруженности (2.3.2) и (2.3.3)

дисперсия минимальна (наибольшее значение σ^2 равно 20.2 при 5.3, в среднем, суперпиров; сеть 41, дуговой алгоритм, 3-я функция разгруженности). Далее приведем данные по среднему числу связанных подграфов суперпиров в сетях (результаты показаны на Рис. 7.) Обозначения те же, что и для Рис. 5.

Представленные результаты показывают преимущество кластеризации с использованием функции (2.3.2) разгруженности узлов сети, т.к. в 67% случаев все выделенные суперпиры образовывали связную сеть. Также стоит

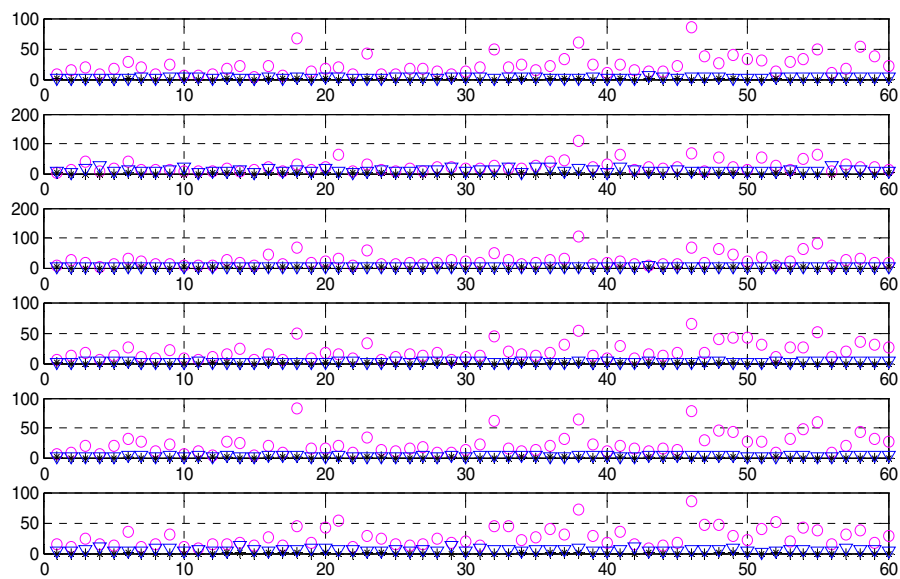


Рис. 6. Дисперсия σ^2 найденного числа суперпиров

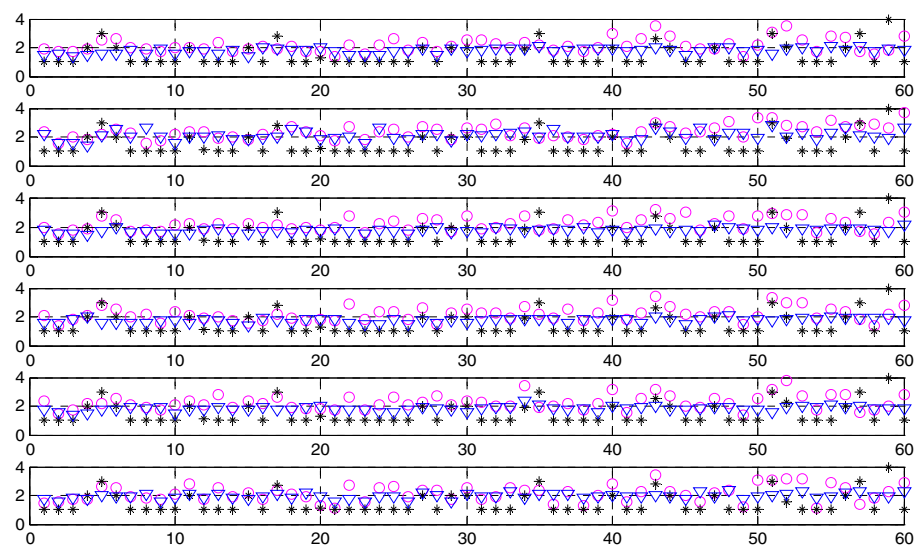


Рис. 7. Среднее число связанных подграфов суперпиров

отметить, что ни один из использованных алгоритмов последовательного заполнения сети качественно не повлиял на финальный результат, поэтому в дальнейшем будут рассматриваться эксперименты с 2 - 3 алгоритмами заполнения сетей. (В дальнейшем, это будут *дуговой*, *субоптимальный минимально разрезный* и *субоптимальный дуговой* алгоритмы).

Приведем результаты **эксперимента 2** для различных алгоритмов последовательного заполнения сетей.

На Рис. 8 – Рис.10 представлены результаты по нахождению доли покрытия исходной сети кластерами узлов при различных значениях радиуса кластера узлов, вокруг найденных суперузлов. По оси абсцисс отложены номера сетей, по оси ординат доля покрытия сетей. Графики (сверху вниз) соответствуют радиусам подграфов кластеров узлов $r=1$, $r=2$, $r=3$ и $r=4$ соответственно.

Во втором эксперименте наиболее удовлетворительным считается результат, при котором

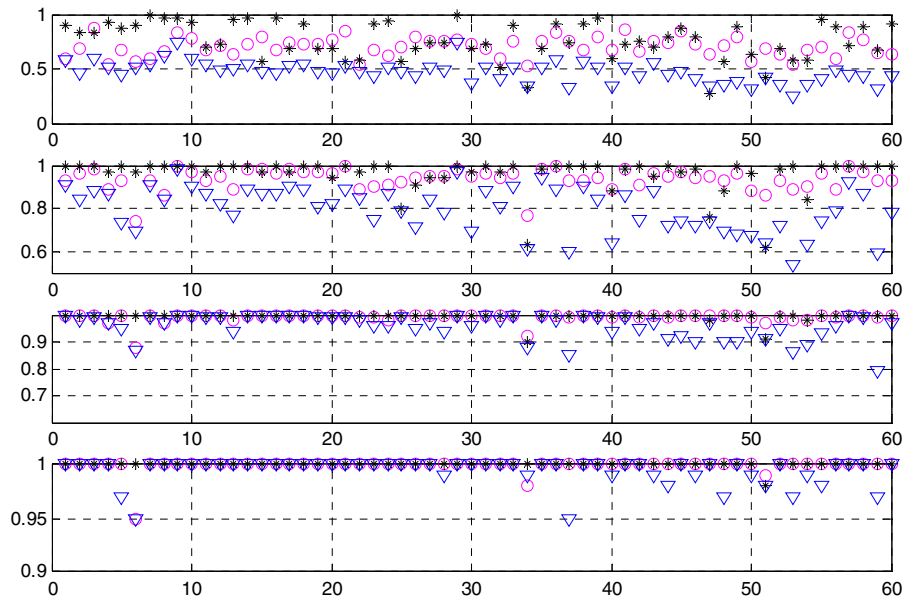


Рис. 8. Доля покрытия сетей кластерами узлов. Дуговой алгоритм

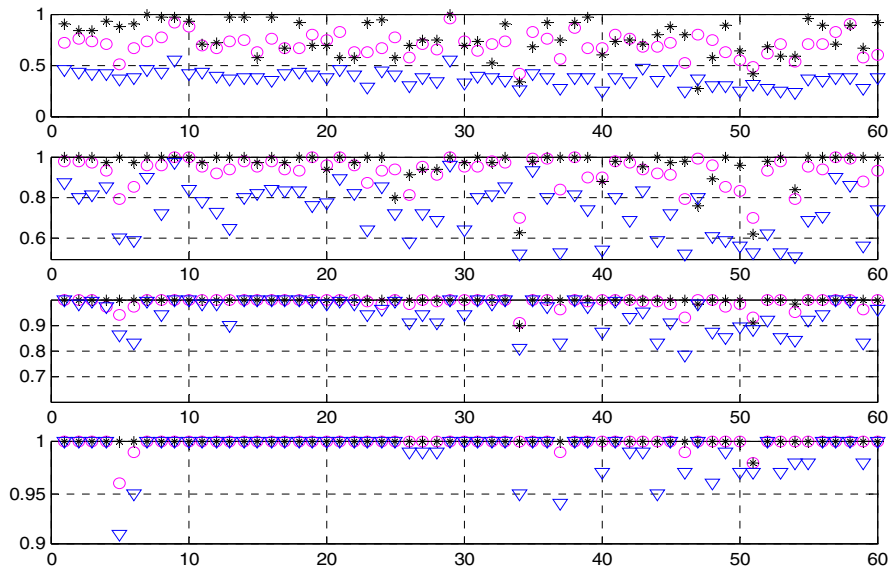


Рис. 9. Доля покрытия сетей кластерами узлов. Субоптимальный минимально разрезный алгоритм

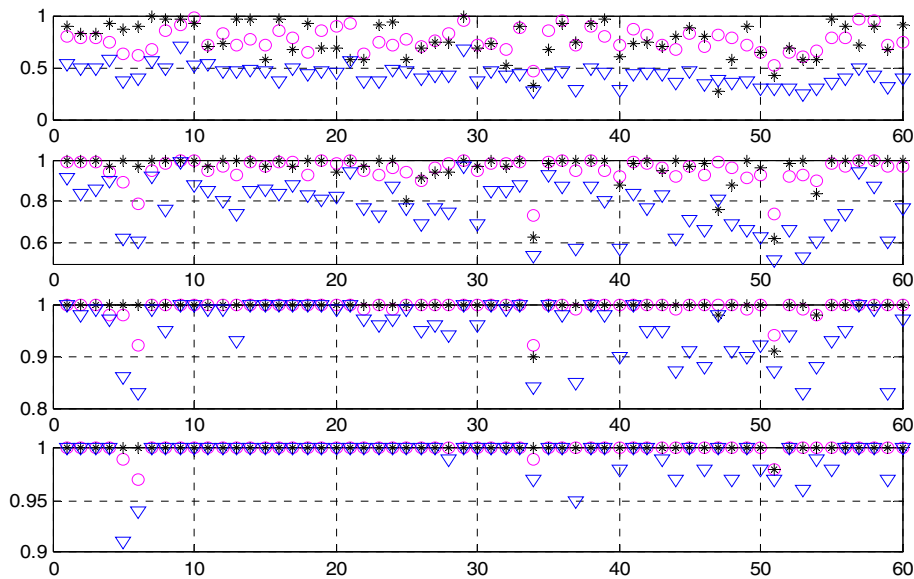


Рис. 10. Доля покрытия сетей кластерами узлов. Субоптимальный дуговой алгоритм

покрытие сети кластерами узлов максимально. Приведем сводную таблицу доли полностью покрытых кластерами сетей для каждой функции разгруженности.

Приведенные рисунки (Рис.8 – Рис.10) показывают, что для кластеризации по функции разгруженности узлов (2.3.2) в случае заполнения сети по дуговому алгоритму, для 60% сетей полное покрытие кластерами сети происходит при радиусе кластера – 2 дуги, для 96% сетей полное покрытие кластерами сети происходит при радиусе кластера – 3 ребра. Для функции разгруженности (1) при радиусах кластеров в 2 и 3 ребра полное покрытие сетей наблюдается у 17% и 85% сетей соответственно. Результаты для кластеризации по функции разгруженности (2.3.3) заметно хуже.

Таким образом, функция разгруженности узлов (2.3.2) лучше всего удовлетворяет требованиям к результатам экспериментов, указанных в пунктах 3 и 4, поскольку кластеризация, основанная на данной функции разгруженности, обладает следующими свойствами:

1. малое число (1-2) связных подграфов суперпиров;
2. малая дисперсия числа выделяемых суперпиров и подграфов суперпиров;
3. большая доля покрытия сети кластерами узлов вокруг суперпиров при малых радиусах кластеров.

Следовательно, вторую функцию разгруженности узлов (2.3.2) можно рекомендовать при построении алгоритмов двухуровневой маршрутизации в сетях передачи данных.

Заключение

В данной работе была рассмотрена задача выделения узлов с дополнительной функциональностью (названных суперпирами/суперузлами) в сетях передачи данных, а также присоединения оставшихся узлов к ним. Была предложена характеристика разгруженности узлов и три функции, определяющие ее, по которой производилось нахождение кандидатов в узлы с дополнительной функциональностью

	Алгоритм	Дуговой	Субоптимальный минимально разрезный	Субоптимальный дуговой
Разгруженность 1	Радиус 2	17%	33%	35%
	Радиус 3	83%	76%	85%
Разгруженность 2	Радиус 2	60%	57%	57%
	Радиус 3	93%	93%	93%
Разгруженность 3	Радиус 2	5%	2%	2%
	Радиус 3	33%	36%	33%

и предложен алгоритм их выделения, основанный на нахождении минимального разреза на графе признаков.

Были проведены численные эксперименты по сравнению результатов нахождения суперпириров при помощи различных функций загрузки узлов. Сравнение проводилось по следующим направлениям:

- нахождение общего числа суперпириров, оценка их среднего числа и дисперсии для каждой сети;
- нахождение числа связных подграфов, образуемых суперпирами;
- при выделении связного подграфа суперпириров и наращивания вокруг них оставшихся узлов, нахождение степени покрытия сети образовавшимися группами узлов.

По результатам проведенных экспериментов для функций загрузки узлов (3.2.2) и (3.2.3) дисперсия числа суперпириров минимальна (максимальное значение дисперсии 4.49 суперпира), когда для первой функции загрузки узлов максимальное значение дисперсии 10.42. Вторая функция загрузки узлов на 67% сетей выделяла суперпиры, объединенные в один связный подграф. В таблице показано, что при наращивании узлов вокруг суперпириров, наилучшие результаты достигались для первой и второй функций загрузки узлов (81.3% и 93% процента полностью покрытых сетей, при радиусе кластера 3 дуги). Таким образом, наилучшей функцией загрузки узлов оказалась функция, основанная на совокупности степени вершины и суммы остаточных пропускных способностей инцидентных дуг (функция 3.2.2).

Предложенный подход разделения узлов позволяет ввести двухуровневую маршрутизацию потоков данных в сети, а именно от узла абонента через локальный выделенный узел, далее через подсеть суперпириров, и оканчивается на узле адресате, который также принадлежит одному из суперпириров. Это позволяет ускорить вычисление маршрута передаваемого потока данных по сравнению с существующими алгоритмами последовательного заполнения сети потока. В практическом приложении введение двухуровневой маршрутизации позволяет уменьшить размер таблиц маршрутизации на

каждом узле ввиду уменьшения числа возможных связей каждого узла.

Литература

1. Сухорослов О.В. Пиринговые системы: концепция, архитектура и направления исследований. / Проблемы вычислений в распределенной среде: прикладные задачи. Труды ИСА РАН. - М.: РОХОС, 2004, с.7-43.
2. Курочкин И.И., Корх А.В. Использование идеологии пиринговых сетей для организации трафика в специализированных телекоммуникационных сетях. / Материалы XVII международной конференции по вычислительной механике и современным прикладным программным системам ВМСППС, 2011. - 832 с. (с.250 - 252)
3. 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
4. J. Ledlie, M. Mitzenmacher, M. Seltzer. Wired Geometric Routing / In Proceedings of IPTPS 2007
5. Frank Dabek, Russ Cox, Frans Kaashoek, Robert Morris. Vivaldi: A decentralized Network Coordinate System. / SIGCOMM'04, Aug. 30-Sept. 3, 2004, Portland, Oregon, USA
6. H.V. Jagadish. Linear Clustering of Objects with Multiple Attributes. / ACM, 1990
7. T. Asano, D. Ranjan, T. Roos, E. Welzl, and P. Widmaier. Space Filling Curves and Their Use in Geometric Data Structures / Theoretical Computer Science, vol. 181, no. 1, pp. 3-15, July 1997.
8. Zhichen Xu, Mallik Mahalingam, Magnus Karlsson. Turning Heterogeneity into an Advantage in Overlay Routing. / IEEE IFOCOM, 2003
9. C.Plaxton, R.Rajaram, A.W.Richa. Accessing nearby copies of replicated objects in a distributed environment. / Proceedings of the Ninth Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA), June 1997
10. 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
11. 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, pages 18-27, 2005.
12. Yuri Boykov, Olga Veksler, Ramin Zabih. Fast Approximate Energy Minimization via Graph Cuts / IEEE Transactions on pattern analysis and machine intelligence, november 2001, V. 23, N. 11
13. Курочкин И.И., Корх А.В. Сегментация изображений посредством разрезов на графах / Теория и практика системного анализа: Труды I Всероссийской научной конференции молодых ученых. - Т. II. - Рыбинск: РГАТА имени П. А. Соловьева, 2010. - 218 с. (с. 168-172)

14. 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.
15. 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.
16. 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.

Гринберг Яков Рувимович. Центр Грид-технологий и распределенных вычислений. Кандидат технических наук. Автор 26 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 33 работ. Область научных интересов: имитационное математическое 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 10 работ. Область научных интересов: математическое моделирование телекоммуникационных сетей, сегментация растровых и спектральных изображений, алгоритмы сжатия видеопотоков. E-mail artkorkh@gmail.com