

АЛГОРИТМ ОПТИМАЛЬНОЙ КЛАСТЕРИЗАЦИИ ДЛЯ СЕТЕЙ СОТОВОЙ СВЯЗИ

Попов В.А., Воропаева В.Я., Верховский Я.М.

Донецкий национальный технический университет, г. Донецк
кафедра автоматики и телекоммуникаций, e-mail: vap@fcita.dn.ua

Abstract. Vladislav Popov, Viktoriya Voropayeva, Yakov Vehchovskiy. Optimal clusterization algorithm for cell networks. The main purpose of this paper is to propose algorithm of clusters optimal localization for the cell network planning. Greedy Algorithm and 2 steps of optimization are describing.

Общая постановка проблемы. При планировании сетей сотовой связи необходимо обеспечить покрытие на некоторой территории и предоставление различных сервисов с высокими показателями качества. Причем, если недавно усилия проектировщиков были направлены на достижение требуемого уровня сигнала в каждой точке области покрытия, то в последнее время на первое место выходят показатели экономической эффективности разворачиваемых сетей. Вследствие этого, аналитические методы планирования, в основе которых лежат энергетические характеристики сети, становятся недостаточно эффективными и должны быть дополнены прогнозированием абонентской нагрузки и возможностью адаптации к изменениям условий спроса на услуги (интенсивности трафика) в каждой точке проектируемой сети. Задачи «планирования по запросу» и разработки алгоритмов стохастической идентификации в системах мобильной связи рассматривались российскими и зарубежными исследователями [1, 2]. В данной работе предлагается алгоритм оптимальной кластеризации, отличающийся возможностью адаптации к изменениям условий спроса на услуги (интенсивности трафика) в каждой точке проектируемой сети.

Постановка задач исследования. На первичном этапе планирования сети сотовой связи оператору необходимо иметь статистические данные о предполагаемом количестве абонентов в зоне покрытия, объеме и наборе услуг для каждого абонентского сегмента, а также данные об интенсивности трафика. Но для получения этих данных нужно провести длительные и дорогостоящие исследования в районе планирования, результаты которых имеют высокую погрешность из-за мобильности абонентов, хаотичности распределения нагрузок, сложности обработки огромного числа статистических данных. Исходя из этого, целесообразно трафик в определенных районах зоны покрытия привязывать к соответствующим точкам – *узлам спроса*, совокупность которых может дать начальное представление о положении базовых станций сети для первичного проектирования. Т.е. узлы спроса представляют некоторую группу пользователей сети и отражают характер трафика, генерируемого данной группой.

Наложив растровую сетку определенной размерности (100×100м, 10×10м и т.д.) на область планирования можно разбить ее на области обслуживания, которые в дальнейшем будут тяготеть к одному из узлов спроса. Набор элементов сетки дает возможность сформировать матрицу трафика, при этом трафик элемента матрицы рассчитывается по формуле:

$$E_{geo}(x,y) = \sum_i z_i \cdot \eta_i(x,y), \quad (1)$$

где η_i – интенсивность трафика в данной точки местности проектирования. Данная величина может быть определена на основании имеющихся статистических данных,

которыми располагают операторы уже развернутых сетей. А функция $\delta_i(x, y)$ принимает значения:

$$\delta_i(x, y) = \begin{cases} 0 & \text{если } i \notin (x, y) \\ 1 & \text{если } i \in (x, y) \end{cases} \quad (2)$$

Таким образом, каждый элемент матрицы формируется в зависимости от плотности населения (трафика) в данной точке и принадлежности той или иной категории пользователей к элементу матрицы.

Сформированная таким образом матрица, элементы которой имеют существенный разброс значений, позволяет выполнить т.н. *кластеризацию* сети. При этом под «кластером» будем понимать область проектирования, содержащую элементы с относительно высокой интенсивностью трафика, отделенную от других областей пространствами с относительно меньшим трафиком.

После кластеризации зоны обслуживания, для каждого кластера необходимо определить положение узла спроса. Узлом спроса является центр тяготения трафика всех входящих в кластер элементов (центр тяжести фигуры описывающей кластер), поэтому его координаты следует определять в соответствии с выражениями:

$$X_{DN} = \frac{1}{\sum_{\forall e \in E(Z)} T_e} \cdot \sum_{\forall e \in E(Z)} x_e \cdot T_e ; \quad (3)$$

$$Y_{DN} = \frac{1}{\sum_{\forall e \in E(Z)} T_e} \cdot \sum_{\forall e \in E(Z)} y_e \cdot T_e , \quad (4)$$

где Z – кластер, для которого определяется узел спроса; e – элемент кластера; T_e – трафик элемента e ; $E(Z)$ – количество элементов кластера; x_e, y_e – географические координаты центров элементов кластера.

Таким образом, точка тяготения $D_N(X_{DN}, Y_{DN})$ определяет позицию узла спроса. Трафик полученных узлов спроса соответствует суммарному трафику элементов матрицы T_Z , входящих в кластер Z .

Очевидно, что полученные узлы могут иметь значения трафика, существенно отличающиеся друг от друга. Однако для оптимального использования аппаратуры базовых станций этот показатель должен быть примерно одинаков для разных узлов. Отсюда вытекает задача объединения узлов, решаемая различными методами [3].

Решение задач и результаты исследования. В работе предлагается использовать известный в теории графов поглощающий алгоритм (Greedy Algorithm), основанный на «поглощении» некоторым начальным K -м узлом (кластером) соседнего N -го узла согласно условиям:

$$T_K + T_N < T_{\max} ; \quad (5)$$

$$A_K + A_N \leq A_{\max} \quad (6)$$

$$T_N \leq T_i \forall i \in Nh(K), \quad (7)$$

где T_K – трафик узла K ; T_N – трафик узла N ;
 A_K и A_N – площади узлов K и N ;
 T_{max} и A_{max} – максимально допустимая нагрузка и площадь покрытия узла спроса
 $Nh(k)$ – число соседей узла K .

Таким образом, среди соседних узлов выбирается узел с наименьшим трафиком. Перед поглощением нового узла проверяется суммарная нагрузка и площадь покрытия по отношению к граничным параметрам алгоритма T_{max} и A_{max} .

При определении этих параметров рекомендуется руководствоваться следующими соображениями. Максимальная площадь A_{max} , которую может покрыть узел спроса, определяется стандартом сети сотовой связи, ограничивающим максимальную площадь покрытия секторной антенны. При выборе данного параметра необходимо учитывать рельеф местности в области установки БС, влияющий на распространения электромагнитных волн, а также временную и пространственную нестабильность. Данная величина будет довольно сильно отличаться в разных районах планирования, но, в целом, можно вывести среднее значение для всей области проектирования, которое будет являться априорной информацией на начальном этапе развертки сети. При эксплуатации сети данный параметр может быть уточнен или изменен на основании данных, полученных при реализации моделей распространения радиоволн (модель Ли, модель Окумура-Хаты и др.)

Для оценки максимального значения трафика T_{max} можно использовать модель Эрланга для систем с отказами [3]:

$$p_a = \frac{\frac{T^n}{n!}}{\sum_{i=0}^n \frac{T^i}{i!}} \quad (8)$$

Данная формула является табулированной. Но, как оказалось на практике, это не всегда удобно для расчетов нагрузки при проектировании сотовой сети. Очевидная сложность процедуры определения нагрузки непосредственно при помощи формулы не позволяет рекомендовать ее для инженерного использования, тем более, что по ее виду ничего нельзя сказать о характере зависимости величины допустимой нагрузки от значений вероятностей отказа и числа каналов.

Точное решение уравнения (8) относительно нагрузки T невозможно. В то же время со сколь угодно высокой точностью оно может быть получено одним из численных методов, например, с помощью итерационной процедуры Ньютона или с использованием формулы Стирлинга.

Решая уравнение (8) относительно параметра нагрузки получаем:

$$T = \begin{cases} n \left[1 - \sqrt{1 - \left(p_a \sqrt{\frac{\pi n}{2}} \right)^{\frac{1}{n}}} \right], \text{ при } p_a \leq \sqrt{\frac{2}{\pi n}}, \\ n + \left[\frac{\pi}{2} + 2 \cdot n \cdot \ln \sqrt{\left(p_a \sqrt{\frac{\pi n}{2}} \right) - \sqrt{\frac{\pi}{2}}} \right], \text{ при } p_a > \sqrt{\frac{2}{\pi n}}, \end{cases} \quad (9)$$

Задав значения вероятности потерь p_a , а также число каналов n можно провести ориентировочную оценку нагрузки T , а в случае алгоритма – трафикового параметра T_{max} .

Если новые параметры удовлетворяют условиям, то производится склеивание элементов матрицы:

$$T_Z = T_K + T_N ; \quad (10)$$

$$A_Z = A_K + A_N ; \quad (11)$$

$$E_Z = E_K \cup E_N \quad (12)$$

где T_Z , A_Z – соответственно трафик и площадь сформированного кластера; E_Z – общее число элементов, входящих в кластер.

После формирования кластера, определяется узел спроса, как центр тяготения на основании выражений (3), (4). В результате работы алгоритма распределение трафика между узлами спроса может иметь вид, показанный на рис. 1.



Рис. 1 – Гистограмма распределения трафика между узлами спроса

Из полученной гистограммы видно, что нагрузка разных областей сети все еще существенно отличается. Для организации равномерного распределения нагрузок между узлами спроса, а значит и базовыми станциями, предлагается провести два уровня оптимизации системы:

- ликвидация узлов с маленькой нагрузкой;
- распределение узлов матрицы по узлам спроса в пределах заданных границ.

На первом этапе оптимизации первичных данных кластеризации следует определить узлы спроса, внутренний трафик которых менее заданной величины T_{min} и распределить их компоненты между соседями. При этом, соседние узлы должны обладать такими параметрами T_{max} и A_{max} , чтобы иметь возможность принять в свой состав новые элементы, не превысив эти граничные значения. Для решения этой проблемы предлагается ввести дополнительный параметр R_{Tmax} , который будет характеризовать максимальное отклонение от T_{max} , при котором система может функционировать без существенных потерь. Параметр R_{Tmax} определяется техническими характеристиками аппаратуры БС и характеризует предельный режим работы оборудования.

При исследованиях были выбраны значения:

$$T_{min} = 0,6 \cdot T_{max}, R_{Tmax} = 1,1-1,3.$$

Алгоритм можно представить с помощью выражений:

$$T_K < T_{\min} ; \quad (13)$$

$$T_N + T_E < R_{T_{\max}} \cdot T_{\max} ; \quad (14)$$

$$A_N + A_E \leq A_{\max} ; \quad (15)$$

$$T_N \leq T_i \forall i \in Nh(K) , \quad (16)$$

где T_K – трафик узла спроса K ; T_N – трафик единичного узла N ; T_E – трафик узла спроса E ; A_E и A_N – площади узлов E и N ; A_{\max} – максимальная допустимая нагрузка площадь покрытия узла спроса; $Nh(k)$ – число соседей узла K .

Таким образом, ход первого этапа оптимизации состоит в следующем. Анализируем полученные после кластеризации узлы спроса на предмет соответствия выражению (13). Если данное условие выполняется, то кластер подлежит разделению между соседями (является донором). Выбираем кластер-акцептор, для присоединения узлов донора в соответствии с условиями (14), (15) и присоединяем их. Параметром выхода алгоритма оптимизации является отсутствие узлов, удовлетворяющих условию (13).

В результате проведения первой фазы оптимизации характер распределения трафика между узлами спроса может иметь вид, как на рисунке 2 а).

Второй этап оптимизации предполагает улучшение распределения нагрузки между сформированными узлами спроса на основании введения дополнительных граничных параметров. Параметрами алгоритма вторичной оптимизации являются:

- Максимальное значение трафика T_{select_max} , которое должен иметь узел спроса, чтобы его элемент можно было передать другому узлу спроса на обслуживание.
- Минимальное значение трафика T_{select_min} , которое должно соответствовать узлу спроса, чтобы он мог присоединить дополнительный элемент матрицы.

Таким образом, величина T_{select_max} определяет кластеры, которые могут стать донорами для узлов спроса с меньшим трафиком. T_{select_min} , в свою очередь, обозначает узлы спроса, готовые на добавление новых элементов. От выбора области значений двух этих параметров зависит степень равномерного распределения трафика в области планирования. Трафик полученных узлов спроса будет варьироваться в области заданных значений. В аналитическом виде вторую фазу оптимизации можно представить следующим образом:

$$T_{select_max} < R_{T_{\max}} \cdot T_{\max} . \quad (17)$$

– условие, при котором узел спроса удовлетворяет оптимизации:

$$T_{select_min} \leq T_K \leq T_{select_max} ; \quad (18)$$

– условия на перекомбинацию:

$$T_K > T_{select_max} ; \quad (19)$$

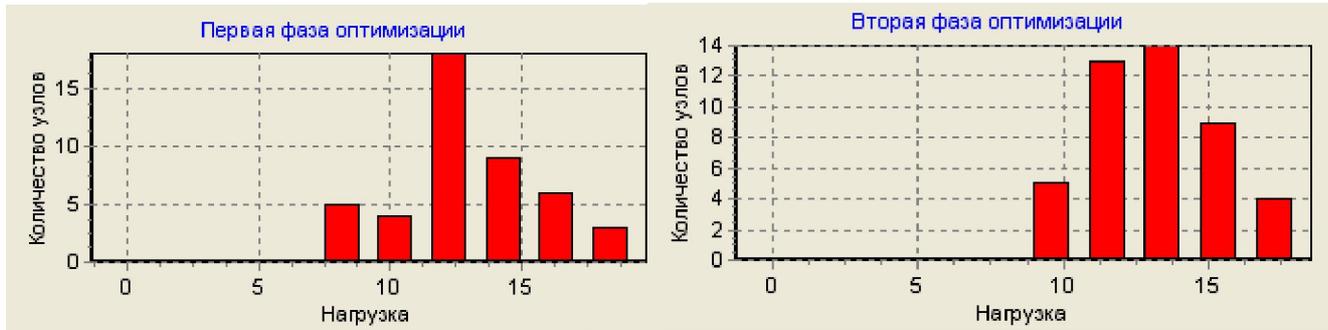
$$T_E < T_{select_min} ; \quad (20)$$

$$T_N + T_E < R_{T_{\max}} \cdot T_{\max} ; \quad (21)$$

$$A_N + A_E \leq A_{\max}; \quad (22)$$

$$T_N \leq T_i \forall i \in Nh(K). \quad (23)$$

В результате выполнения второй фазы оптимизации характер распределения трафика между узлами спроса может иметь вид, как на рисунке 2 б).



а)

б)

Рисунок 2– Гистограмма распределения трафика после первой (а) и второй (б) фазы оптимизации

Выводы:

1. Разработанный алгоритм оптимальной кластеризации обеспечивает равномерное распределение нагрузки между базовыми станциями, что позволит эффективно использовать всю имеющуюся информационную емкость.
2. Алгоритм позволяет прогнозировать зоны покрытия базовых станций в форме окружности, охватывающей зону полученного кластера. Радиус окружности будет зависеть от допустимой площади покрытия и допустимого трафика для БС сети.
3. Предложенный алгоритм отличается возможностью адаптации к изменениям условий спроса на услуги (интенсивности трафика) в каждой точке проектируемой сети.

Литература.

1. Кузнецов И.В., Блохин В.В., Султанов А.Х. Разработка модели и алгоритмов стохастической идентификации ситуации в системе мобильной связи.// Материалы Шестой Международной научно-технической конференции “Проблемы техники и технологии телекоммуникаций”. Уфа, 28 – 30 ноября 2005 г. с. 141-148.
2. Tutschku K. Demand-based Radio Network Planning of Cellular Mobile Communication Systems.: In Proceedings of the IEEE Infocom 2000, San Francisco, USA.
3. М. Ван Стеен , Э. Таненбаум. Распределенные системы. Принципы и парадигмы. :Пер. с англ. – СПб.:Издательство “Питер”, 2003. –880 с.