СВЯЗЬ

Математическая модель доступа к FMC-услугам с учетом равномерной балансировки нагрузки между сетями доступа при миграции абонента

Mathematical model of access to FMC-services considering to uniform distribution of traffic load between access networks and subscriber and subscriber migration

Ключевые слова: сети доступа – access networks; нагрузка – load; трафик – traffic; алгоритм – algorigthm; повышение эффективности – increase efficiency; оптимизация – optimization.

В статье представлена математическая модель доступа к конвергентным услугам, адаптируемая в согласии с сетевыми конфигурациями, через гетерогенные сети доступа в рамках концепции Always Best Connected (ABC). Приведены алгоритмы равномерного распределения и балансировки нагрузки между сетями доступа, позволяющие эффективно использовать сетевой ресурс.

The article represents mathematical model of adaptive access to convergent services considering to network configurations via heterogeneous access networks in case of Always Best Connected (ABC) concept. Algorithms of uniform traffic load distribution and load balancing between access networks are presented and used to effectively utilize network resource.

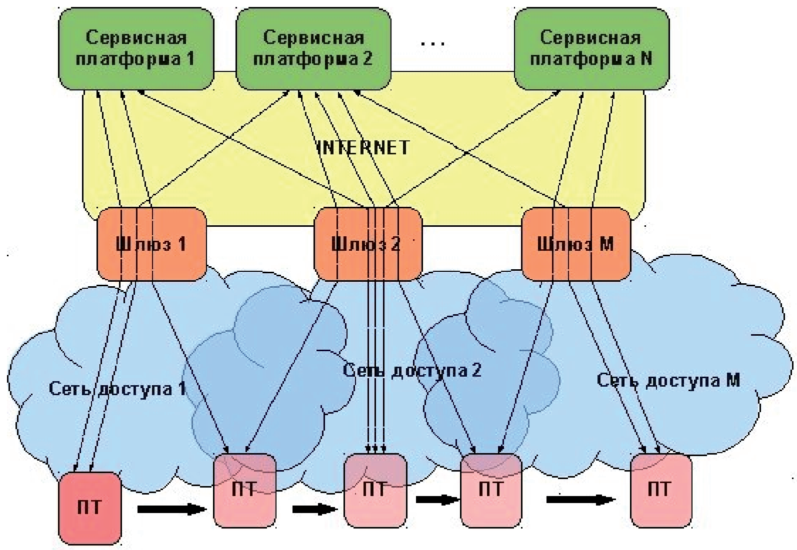
САДОВНИКОВ / SADOVNIKOV V.
Владимир Юрьевич
(sadko_sobaka@rambler.ru)
аспирант кафедры систем коммутации и распределения информации Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург

Происходящая в области телекоммуникаций конвергенция фиксированной и мобильной связи (FMC – Fixed-Mobile Convergence) призвана объединить существующее множество гетерогенных сетей, порой несовместимых между собой и представляющих различные услуги, в единое целое – мультисервисную сеть.

Попытка создания мультисервисной сети на базе ISDN потерпела неудачу в связи с ограниченностью ресурса и дороговизной реализации, но по мере развития сети ОКС № 7 возникла идея доступа к инфокоммуникационным услугам, адаптируемого в согласии с сетевыми конфигурациями. Появившаяся необходимость объединения услуг в рамках мультисервисных услуг привела к концепции NGN (NGN – Next Generation Networks) и вытекающему из нее новому стандарту IMS (IP Multimedia Subsystem). Однако вопрос доступа к услуге остается открытым – абонент ограничен в перемещении. Сегодня современной полномасштабной трактовкой такого доступа является концепция Always Best Connected (ABC), которая подразумевает автоматический выбор оптимального для пользовательского устройства типа соединения из доступных [1].

Оптимальность подразумевает гладкое (без заметного разрыва связи) переключение между гетерогенными сетями доступа, прозрачную регистрацию абонентских терминалов в сетях доступа, а также единую биллинговую систему [2]. Предлагаемые методики взаимодействия с сетями доступа базируются на следующих критериях: общей пропускной способности сетей (повышение эффективности выделения ресурсов сетей) и требованиях к трафику, определяемых пользовательскими приложениями (максимизируется общая емкость системы), стоимостных характеристиках ресурсов сетей (оптимизация финансовых затрат), энергозатратах (экономия энергии мобильными устройствами) [3–6].

Оптимизация по одному или нескольким из перечисленных критериев сводится исследователями к усложнению логики принятия решения о выделении ресурса на базе алгоритма First Fit



Доступ пользовательского терминала к конвергентным услугам

Decreasing (FFD) [7]. Однако эксперименты показывают, что предлагаемые варианты алгоритмов, выдающие наиболее близкие к оптимальным результаты по энергопотреблению и пользовательским предпочтениям, не являются лучшим решением, так как не всегда запрошенная полоса пропускания может быть использована эффективно, т.е. полностью.

Для повышения эффективности использования сетевых ресурсов надо рассматривать трафик системы не с точки зрения ширины полосы пропускания, запрашиваемой услугой, а с точки зрения нагрузки, создаваемой этой услугой. Остается нерешенным вопрос о миграции абонента между сетями доступа и балансировке нагрузки на сети доступа, что немаловажно для Always Best Connected.

На рисунке демонстрирует доступ пользовательского терминала (ПТ) к конвергентным услугам в процессе смены географического местоположения. В любой географической точке можно выделить M сетей доступа и N классов услуг. Нагрузку, обслуживаемую сетями доступа, можно выразить вектором $\vec{y}=\{y_1,\dots,y_m\}$, где $m=1\dots M$.

Основные требования к оценочной функции $o(y)$ сети доступа, позволяющей дать качественную оценку изменения загруженности ресурса каждой сети в момент запроса нового ресурса:

– при отсутствии нагрузки функция должна возвращаться в максимальное значение;

– при максимальной нагрузке на сеть функция возвращается в нулевое значение;

– функция должна быть непрерывной: $\lim\limits_{y\to a} o(y)=o(a)\ \forall a\in(0,1)$;

– функция должна быть монотонно убывающей: $\frac{do(y)}{dy}<0\ \forall y\in(0,1)$;

– функция должна быть выпуклой:

$$\frac{d^2o(y)}{dy^2}<0\ \forall y\in(0,1).$$

Такая оценочная функция позволяет сделать следующие выводы:

1. Менее нагруженные сети доступа имеют больший приоритет по сравнению с более нагруженными сетями.

2. Так как выделение ресурса в более загруженной сети обходится дороже, чем в любой менее загруженной сети, распределение поступающей нагрузки между сетями доступа должно стремиться к равномерному: $\sum\limits_{\substack{i=1\dots M-1,\\ j=i+1\dots M}} y_i - y_j \to min$.

Введем вектор принятия решения о распределении запрашиваемого ресурса между сетями: $\vec{d}=\{d_1,\dots,d_m\}$: $\sum\limits_{m=1}^{M} d_m=1,\ d_m\ge 0$. Вектор $\vec{d}$ всегда имеет единственную отличную от нуля составляющую, которая показывает, в какой сети следует выделить запрашиваемый ресурс. Общую эффективность распределения нагрузки между сетями в момент запроса ресурса можно охарактеризовать следующей функцией:

$$\bar{O}(\vec{d})=\sum_{m=1}^{M}\tilde{o}(y_m,d_m),\ \text{где}\ \tilde{o}(y_m,d_m)=o(y_m+\Delta y_m\cdot d_m).$$

На основании вышесказанного условие, по которому следует выбирать сеть доступа, чтобы максимально сохранить равномерность распределения нагрузки, следующее: необходимо найти такой вектор $\vec{d}$, что $\bar{O}(\vec{d}) \Rightarrow max$. Условие относится к классу задач выпуклого программирования. Следуя основной теореме математического программирования (ОТМП) и методам, представленным в [8, 9], для получения оптимального вектора распределения запрашиваемого ресурса между сетями можно получить функцию Лагранжа для решения задачи:

$$\Phi(\vec{d},\vec{\Psi}) = \bar{O}(\vec{d}) + \Psi\cdot\left(1-\sum_{m=1}^{M} d_m\right).$$

При этом, согласно ОТМП, необходимые и достаточные условия оптимальности будут следующими:

$$\frac{\partial\Phi(\vec{d},\vec{\Psi})}{\partial d_m} = \frac{\partial\bar{o}(y_m, d_m)}{\partial d_m} - \Psi\begin{cases}\le 0, & d_m = 0\\ = 0, & d_m > 0\end{cases}.$$

Учитывая сформулированные требования к оценочной функции в ходе решения дифференциальных уравнений, получаем предварительную оценочную функцию сети доступа: $o(y) = 1 - y^2$. При анализе производных этой функции в граничных точках интервала [0; 1] выяснилось, что в точке «0» функция имеет экстремум, а ближе к точке «1» убывает с постоянной скоростью. В идеале значение ее производной в точке «1» должно стремиться к отрицательной бесконечности, что можно трактовать как невозможность выделения какого-либо ресурса полностью загруженной сетью доступа. Следует проанализировать значения производных схожей функции $o(y) = \sqrt{1-y^2}$:

$$\lim_{y\to 0+0}\frac{do(y)}{dy} = \lim_{y\to 0+0}\left(-\frac{y}{\sqrt{1-y^2}}\right) =$$

$$= 0 \quad \lim_{y\to 1-0}\frac{do(y)}{dy} = \lim_{y\to 1-0}\left(-\frac{1}{\sqrt{(1-y^2)^3}}\right) = -\infty.$$

Это означает, что в точке $y=0$ функция также имеет экстремум, а ближе к точке $y=1$ скорость убывания функции стремится к бесконечности, что вполне подходит к нашим требованиям.

Функция Лагранжа с использованием полученной оценочной функции:

$$\Phi(\vec{d},\vec{\Psi}) = \sum_{m=1}^{M}\sqrt{1-\left(y+d_m\cdot\Delta y\right)^2} + \Psi\cdot\left(1-\sum_{m=1}^{M} d_m\right).$$

Необходимые и достаточные условия оптимальности, согласно ОТМП, будут следующими:

$$\frac{\partial\Phi(\vec{d},\vec{\Psi})}{\partial d_m} = \frac{-\Delta y(y+d_m\Delta y)}{\sqrt{1-(y+d_m\cdot\Delta y)^2}} - \Psi\begin{cases}\le 0, & d_m = 0\\ = 0, & d_m > 0\end{cases}.$$

Полученная функция Лагранжа не требует дальнейшего детального исследования, так как в нашем случае для выбора сети доступа нужно найти одну оптимальную компоненту вектора $\vec{d}$, отличную от нуля, т.е. оценить убывание эффективности распределения нагрузки между сетями доступа при выборе обслуживающей сети доступа:

$$k:\ \min_k\left\{\Delta o_k = \sqrt{1-y_k^2} - \sqrt{1-(y_k+\Delta y)^2}\ \forall k = 1\ldots M\right\}.$$

Найдя k, положим $d_m = \begin{cases}0, & m \ne k\\ 1, & m = k\end{cases}.$

Балансировка нагрузки заключается в правильном перераспределении ресурсов между сетями доступа, что приводит к равномерному распределению нагрузки между ними. Имея информацию о текущем распределении ресурсов, предоставляемых сетями доступа, и информацию об оптимальном распределении ресурсов, можно определить, какая сеть и в каком объеме должна взять на себя дополнительную нагрузку, а какая – освободиться от нее. Информация о текущем распределении ресурсов складывается по мере запроса ресурсов сетей доступа клиентами. Оптимальное распределение необходимо рассчитывать при каждом запросе или освобождении ресурса.

На основании ранее сформулированного принятия решения о выборе сети доступа с целью максимального сохранения равномерного распределения нагрузки решим задачу поиска оптимального распределения ресурсов между сетями доступа, т.е. найдем матрицу распределения занятого между сетями доступа ресурса d, распределив компоненты вектора запрашиваемого ресурса $\vec{v} = \{v_1, \ldots, v_m\}$ таким образом, что $O(\dot{d}) \Rightarrow max$:

$$O(\dot{d}) = \sum_{n=1}^{N} o(\vec{d}^n) = \sum_{n=1}^{N}\sqrt{1-\left(\sum_{m=1}^{M}\frac{\lambda^m_n}{\mu^m_n}\cdot d^m_n\right)^2},$$

Общее количество запрашиваемых услуг, которое необходимо распределить между сетями доступа:

$\sum_{m=1}^{M} v_m = V$. В результате распределения получается матрица d, удовлетворяющая следующим условиям:

$$\begin{cases}\sum_{n=1}^{N}\sum_{m=1}^{M} d^m{}_n = V\\ d^m{}_n \ge 0 \quad \forall m,\, n.\end{cases}$$

Так как оценочная функция сети доступа соответствует всем требованиям выпуклого программирования, сумма оценочных функций по каждой сети также соответствует всем требованиям ОТМП.

Решение находится алгоритмическим способом – методом максимального элемента (ММЭ) [9], в основе которого лежат два основополагающих факта: ресурсы распределяются пошагово, по одному, начиная с нулевого состояния, когда ни на одну сеть не выделено еще ни одного ресурса; на каждом шаге ресурс выделяется на ту сеть, у которой достигается минимальная потеря эффективности – это позволяет найти приближенное к оптимальному решение.

Алгоритм приближенного целочисленного распределения услуг по сетям связи с использованием ММЭ:

1. Получить исходные данные: N, M, вектор $\vec{v}$, матрицы $\dot{\lambda}$, $\dot{\mu}$.
2. Вычислить компоненты нагрузочной матрицы $\Delta\dot{Y}$ для всех сетей и услуг:

$$\Delta Y^m{}_n = \frac{\lambda^m{}_n}{\mu^m{}_n} \quad \forall n=1\ldots N,\ m=1\ldots M.$$

3. Проинициализировать матрицу распределения ресурсов между сетями и услугами: $d_n{}^m = 0 \quad \forall n=1\ldots N,\ m=1\ldots M.$
4. Проинициализировать первоначальные значения оценочной функции для каждой сети доступа: $o_m = 1 \quad \forall m = 1\ldots M.$
5. Инициализировать циклическую переменную: $s=1$.
6. Если компонента вектора запрашиваемых ресурсов равна нулю ($v_s = 0$), перейти к п. 11.
7. Вычислить оценочную функцию для всех сетей доступа с учетом занятия их дополнительной услугой:

$$o^+_m = \sqrt{1-\left(\sum_{n=1}^{N} \Delta Y^m{}_n \cdot d^m{}_n + \Delta Y_s{}^m\right)^2} \quad \forall m = 1\ldots M.$$

8. Найти для $\max_k\{o^+{}_k\} \quad \forall k = 1\ldots M.$
9. Вычислить $d^k{}_n = d^k{}_n + 1$, $v_s = v_s - 1$, присвоить $o_k = o^+{}_k$.
10. Перейти к п. 6.
11. Если $s=N$, перейти к п. 14.
12. Вычислить $s=s+1$.
13. Перейти к п. 6.
14. Вывести результирующую матрицу $\dot{d}$.

Алгоритм перераспределения услуг по сетям связи:

1. Получить исходные данные: M, N, матрицу $\dot{d}$, матрицы, $\dot{\lambda}$, $\dot{\mu}$.
2. Вычислить компоненты нагрузочной матрицы $\Delta\dot{Y}$ для всех сетей и услуг:

$$\Delta Y^m{}_n = \frac{\lambda^m{}_n}{\mu^m{}_n} \quad \forall n=1\ldots N,\ m=1\ldots M.$$

3. Инициализировать переменную: $k=1$.
4. Если $k \ge M-1$, перейти к п. 11.
5. Вычислить

$$Y_r{}^s = \begin{cases}\sum_{n=1}^{N} \Delta Y^s_n \cdot d^s_n + \Delta Y^s_{(r-1)\,mod\,N+1} - \Delta Y^s_{(r-1)\div N+1}, \\ -1 \quad, d^s_{(r-1)\div N+1} \le 0\end{cases}$$

$$d^s_{(r-1)\div N+1} > 0 \qquad \forall r = 1\ldots N\cdot M,\ s = k\ldots M.$$

6. Вычислить

$$dO^t_r = \begin{cases}\sqrt{1-(Y^k_r)^2} + \sqrt{1-(Y^{k+t}_r)^2} - \sqrt{1-\left(\sum_{n=1}^{N}\Delta Y^k_n \cdot d^k_n\right)^2} - \\ -1 \quad, y^k_r \le 0 \vee y^{k+t}_r \le 0\end{cases}$$

$$-\sqrt{1-\left(\sum_{n=1}^{N}\Delta Y^{k+t}_n \cdot d^{k+t}_n\right)^2}\ ,\ y^k_r > 0 \wedge y^{k+t}_r > 0.$$

7. Найти r,t для

$$\forall r = 1\ldots N\cdot M,\ t = 0\ldots N-k-1.$$

Если $\max_{k,r}\{dO^t_r\} \ge 0$ перейти к п. 12.

8. Вычислить

$$\forall r = 1\ldots N\cdot M,\ t = 0\ldots N-k-1.$$

9. Вычислить $d^{m_1}_{n_1} = d^{m_1}_{n_1} - 1,\ \ d^{m_1}_{n_2} = d^{m_1}_{n_2} + 1,$

$$d^{m_2}_{n_2} = d^{m_2}_{n_2} - 1,\ \ d^{m_2}_{n_1} = d^{m_2}_{n_1} + 1.$$

10. Перейти к п. 5.
11. Вычислить $k=k+1$.
12. Перейти к п. 5.
13. Вывести результирующую матрицу $\dot{d}$.

Первый алгоритм позволяет распределить услуги между сетями доступа, получив приближенное оптимальное решение задачи равномерного распределения нагрузки между сетями. Второй алгоритм, учитывая последовательность распределения нагрузки по классам услуг (сначала распределяются услуги первого класса, затем – второго и т. д.), дополнительно оптимизирует полученное с помощью первого алгоритма решение, перераспределяя услуги по сетям связи. В следующей статье представлены результаты распределения ресурсов сетей доступа, полученные данными алгоритмами, по сравнению с алгоритмами, предложенными в [7].

Литература

1. *Рудометов Е., Рудометов В.* Мобильный мир в будущем. – www.fcenter.ru.
2. *Rajesh Mishra.* Always best connected architecture and design. – Ericsson berkley wireless center.
3. *Furuskar A.* Allocation of multiple services in multi-access wireless networks. – Mobile and wireless communication networks. September 2002. – P. 261–265.
4. *Furuskar A.* Radio resource sharing and bearer service allocation for multi-bearer service, multi-access wireless networks – methods to improve capacity. – www.media.lib.kth.
5. *Furuskar A., Zander J.* Multi-service allocation for multi-access wireless systems. – Transactions on Wireless Communication.
6. *Fodor G., Furuskar A., Lundsjo J.* On access selection techniques in always best connected networks. – Specialist seminar on performance evaluation of wireless and mobile systems. – August 2004.
7. *Bo Xing, Nalini Venkatausbramanian.* Multi-constraint dynamic access selection in always best connected networks. – Donald bren school of information and computer sciencies. – Univercity of California.
8. *Дымарский Я.С., Крутякова Н.П., Яновский Г.Г.* Управление сетями связи: принципы, протоколы, прикладные задачи. – М.: ИТЦ «Мобильные коммуникации», 2003.
9. *Дымарский Я.С.* Методы оптимизации сетей связи. – СПб.: Изд-во СПбГУТ, 2003.