

УДК 621.395

МОДЕЛЬ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ РЕСУРСАМИ СЕТИ NGN С УЧЕТОМ ТРЕБОВАНИЙ SLA

С.Ш. Кутбитдинов, заместитель директора ГУП «UNICON.UZ»; kutbitdinov@unicon.uz

Ключевые слова: NGN, алгоритм дифференцированных услуг DiffServ, сетевые ресурсы, SLA, штрафные санкции, функция полезности, оптимальный план распределения ресурсов.

В настоящее время общепризнанной основой для дальнейшего развития сетей связи общего пользования (ССОП) является концепция сетей связи следующего поколения – NGN (Next Generation Network), реализуемая на базе различных технических решений, прежде всего Softswitch [1–5] и IMS (IP Multimedia Subsystem) [6, 7]. Как известно, концепция NGN предполагает создание сети связи с гарантированным уровнем качества обслуживания пользователей, что достигается путем создания новых механизмов управления качеством обслуживания и установления определенных взаимоотношений между операторами связи, а также между оператором связи и пользователем на основе заключаемых соглашений об уровне обслуживания – SLA (Service Level Agreement) [8–10]. В соответствии с ITU-T Y.1291 [11] для обеспечения гарантированного уровня качества обслуживания в сетях NGN в качестве базового рекомендован алгоритм дифференцированных услуг (DiffServ).

Стратегическая цель управления сетевыми ресурсами NGN. Принципиально новый подход к обеспечению гарантированного уровня качества обслуживания пользователей в сети NGN требует, соответственно, и нового подхода к распределению ее сетевых ресурсов. Под сетевыми ресурсами NGN здесь и далее будем понимать в первую очередь производительность систем распределения информации, например совокупности серверов, образующих платформу IMS, а также пропускную способность сети в целом и отдельных ее сегментов, в частности сетей доступа.

Как показано в [1, 4], функциональность сети NGN позволяет реализовать различные стратегии управления сетевыми ресурсами, в основу которых должна быть положена идея оптимального функционирования сети связи. Под оп-

тимальным функционированием сети NGN будем понимать достижение некоей стратегической цели управления сетевыми ресурсами посредством решения следующих задач:

- обеспечение качества предоставления услуг в соответствии с нормами, определенными SLA;
- повышение эффективности использования сетевых ресурсов;
- обеспечение максимальной экономической эффективности сети связи.

В данной статье предпринята попытка разработать подход к выбору стратегической цели управления ресурсами сети NGN, в котором был бы учтен положительный эффект от предоставления услуг и обеспечения заданного уровня их качества и отрицательный – от снижения качества предоставляемых услуг.

Нормы SLA и штрафные санкции. В соответствии с [8–10] в SLA должны быть определены нормы на основные показатели качества услуг, предоставляемых оператором связи, и описаны методы их контроля. В общем случае SLA может содержать также определение штрафных санкций, применяемых к оператору связи при предоставлении им услуг с уровнем качества ниже установленного данным соглашением. Таким образом, в условиях действия SLA оператор связи потенциально может нести некие расходы, связанные с нарушением заключенного соглашения о качестве предоставляемых им услуг.

Следует заметить, что, если в SLA определены штрафные санкции, действующие только при снижении качества, оператор связи несет прямые расходы, связанные с выплатой штрафов. При отсутствии же штрафных санкций потери оператора от снижения качества предоставления услуг в рыночных условиях можно выразить через косвенные расходы, определяемые снижением привлекательности услуг данного оператора для потенциальных пользователей.

Как известно, снижение качества обслуживания пользователей может происходить по ряду причин, зависящих от оператора связи. В данной работе рассматриваются задачи нехватки сетевых ресурсов, вызванной ростом пользовательского трафика.

Распределение сетевых ресурсов. В условиях, когда состав пользователей сети NGN однороден, средний доход оператора связи DI с учетом применяемых штрафных санкций, предусмотренных SLA, будет определяться как

$$DI = DS - DP; DS = nc; DP = lnp, \tag{1}$$

где DS – средний доход оператора связи от предоставления пользователям услуг без учета штрафных санкций; DP – расходы (потери дохода) оператора связи из-за применяемых штрафных санкций, предусмотренных SLA; n – число пользователей сети; c – средний доход оператора связи от предоставления услуг в расчете на одного пользователя (ARPU); l – доля пользователей сети, обслуживаемых с пониженным качеством; p – размер штрафа, предусмотренного за нарушение норм SLA.

Как видно из (1), размер среднего дохода (DI) оператора связи напрямую зависит от доли пользователей, обслуживаемых с пониженным качеством. Экономическая эффективность сети связи E при этом будет определяться разницей между доходом D_1 оператора с сохранением настоящего уровня качества и доходом D_2 оператора при увеличении сетевых ресурсов для повышения качества обслуживания пользователей сети:

$$E = D_1 - D_2, \tag{2}$$

где

$$D_1 = DS - DE - DK; \tag{3}$$

$$D_2 = DS - DP - DE. \tag{4}$$

Здесь DK – капитальные затраты сети; DE – эксплуатационные расходы оператора связи.

С учетом (3) и (4) экономическая эффективность сети E будет определяться как

$$E = -DK + DP, \tag{5}$$

а в расчете на одного пользователя:

$$E_0 = lp - DK_0, \tag{6}$$

где DK_0 – капитальные затраты и эксплуатационные расходы оператора связи в расчете на одного пользователя.

В случае, когда пользовательская база сети неоднородна (состоит из нескольких категорий пользователей, которым предоставляются услуги разного уровня качества), снижение качества обслуживания в различной степени влияет на эффективность функционирования сети связи. Поскольку разные категории пользователей приносят оператору связи разный доход и его издержки при нарушении нормативных показателей качества тоже различны, в целях получения максимального дохода при ограниченном объеме сетевых ресурсов будет логично рационально распределить эти ресурсы между категориями пользователей.

Необходимо отметить, что данную ситуацию можно рассматривать не только с точки зрения снижения уровня качества по сравнению с определенными в SLA нормами, что может, например, иметь место при перегрузке сети связи. При соблюдении норм SLA целью рационального распределения сетевых ресурсов может стать оптимизация косвенных доходов от повышения привлекательности оператора связи в условиях рыночных отношений.

Функция полезности. Рассмотрим случай, когда ССОП обслуживает n пользователей k категорий, т.е. $n_1, n_2, \dots, n_i, \dots, n_k$. Для предоставления услуг n_i пользователям i -й категории требуется сетевой ресурс r_i , за нарушение качества его обслуживания выплачивается штраф размером p_i . Общий объем сетевых ресурсов равен R . Для обслуживания одного пользователя категории i требуется ресурс $r_i^{(0)}$.

При выполнении условия $\sum_{i=1}^k r_i = \sum_{i=1}^k n_i r_i^{(0)} \leq R$ все пользователи получают услуги с надлежащим уровнем качества и штрафные санкции не налагаются (равны нулю). Если $\sum_{i=1}^k r_i > R$, возникает нехватка сетевых ресурсов в количестве $\Delta R = \sum_{i=1}^k r_i - R$ и, как следствие, ухудшается качество обслуживания части пользователей (становится ниже установленного в SLA уровня). В этом случае оператор связи несет расходы, связанные со штрафными санкциями, предусмотренными в SLA.

Распределение сетевых ресурсов состоит в следующем: имеется определенное количество сетевого ресурса R , которое можно использовать k различными способами. Если количество сетевого ресурса, реализуемое i -м способом, обозначить через r_i , то каждому способу можно будет сопоставить функцию полезности $\varphi_i(r_i)$, выражающую доход оператора связи от применения этого способа. Предполагается, что все доходы измеряются в одних единицах, а общий доход равен сумме доходов, полученных от использования каждого способа.

Сформулируем задачу в математической форме – найти общий доход оператора связи от использования сетевых ресурсов всеми способами:

$$\max \{ \varphi_1(r_1) + \varphi_2(r_2) + \dots + \varphi_k(r_k) \} \tag{7}$$

при условии:

1) $r_1 \geq 0, \dots, r_k \geq 0$ (выделяемое количество сетевых ресурсов неотрицательно);

2) $r_1 + r_2 + \dots + r_k = R$ (общее количество сетевых ресурсов равно x).

В качестве функции полезности выберем функцию $\varphi_i(x_i)$, приведенную на рис. 1:

$$\varphi_i(x_i) = \begin{cases} 0 & r_i < 0; \\ \frac{r_i}{r_i^{(0)}} c_i - \left(n_i - \frac{r_i}{r_i^{(0)}} \right) p_i & 0 \leq r_i \leq n_i r_i^{(0)}; \\ n_i c_i & r_i > n_i r_i^{(0)}. \end{cases} \tag{8}$$

Как видно из рис. 1, при отрицательных значениях r_i значение функции $\varphi_i(x_i)$ равно нулю, при значениях r_i от нуля до $n_i r_i^{(0)}$ величина рассматриваемой функции полезности линей-

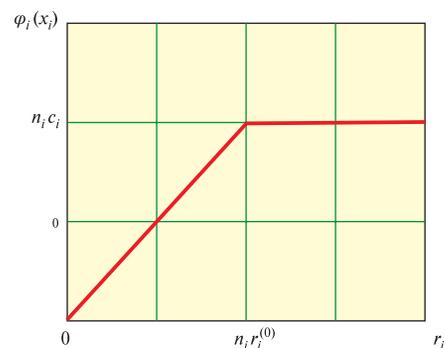


Рис. 1

но растет, причем рост ее прямо пропорционален величине сетевого ресурса и обратно пропорционален размеру штрафных санкций, применяемых к оператору из-за нарушений условий SLA, связанных с недостатком сетевых ресурсов. Если выделяемые сетевые ресурсы равны нулю, функция $\varphi_i(x_i)$ отрицательна и равна величине штрафных санкций $-n_i p_i$. Объем сетевых ресурсов в количестве $n_i r_i^{(0)}$ соответствует потребностям пользователей, и значение функции $\varphi_i(x_i)$ равно $n_i c_i$, штрафные санкции при этом отсутствуют. При большем объеме сетевых ресурсов значение функции полезности не увеличивается.

Используя метод динамического программирования [12], можно найти максимум выражения (7). Ему будет соответствовать вектор оптимального распределения ресурсов всех категорий пользователей $\vec{r} = \{r_1, r_2, \dots, r_k\}$.

Пример расчета оптимальных планов распределения ресурсов сети NGN. В качестве примера рассмотрим ССОП, обслуживающую 10 тыс. пользователей пяти категорий. Для обслуживания одного пользователя требуется сетевой ресурс $r^{(0)}$, средний доход, приносимый каждым пользователем, равен c , штрафные санкции равны p (см. таблицу).

Для обслуживания всех пользователей в соответствии с установленными нормами качества требуется объем сетевых ресурсов в количестве $\sum_{i=1}^k n_i r_i^{(0)} = 2\,850\,000$ условных единиц.

Предположим, что имеет место нехватка сетевых ресурсов в количестве 350 000 условных единиц. Тогда, используя данные таблицы и метод динамического программирования для поиска максимума выражения (7), получаем оптимальный план распределения сетевых ресурсов r (рис. 2).

Доли выделяемых ресурсов для различных категорий пользователей в зависимости от требуемой величины приведены на рис. 3.

Метод динамического программирования позволяет получить оптимальный план распределения сетевых ресурсов, но следует отметить, что в общем случае этот план не единственный. Получение того или иного решения зависит от начальных условий и набора ограничений. Например, при введении в задачу дополнительного ограничения (не менее 80%) на долю сетевых ресурсов от требуемой величины их оптимальное распределение изменится (рис. 4).

Как было показано выше, в качестве сетевого ресурса может выступать также, например, пропускная способность сети связи (скорость передачи данных). В настоящей работе функция полезности рассматривается как кусочно-линейная. Для описания влияния некоторых параметров, характеризующих качество обслуживания пользовательского трафика, могут быть использованы другие функции. Основными параметрами, характеризующими качество обслуживания трафика в сети с пакетной коммутацией, являются временные характеристики доставки данных, такие как задержка и джиттер. Задержки доставки пакета данных от источника до точки назначения определяются временем его передачи на участках транспортной сети и зависят от их пропускной способности и интенсивности обслуживаемого трафика.

| i | $r^{(0)}$ | c | p | n | $\bar{r}, \%$ |
|-----|-----------|------|-----|------|---------------|
| 1 | 1000 | 200 | 100 | 6000 | 40 |
| 2 | 2000 | 300 | 150 | 1800 | 18 |
| 3 | 3000 | 500 | 200 | 1200 | 18 |
| 4 | 4000 | 750 | 250 | 650 | 14 |
| 5 | 5000 | 1000 | 300 | 350 | 10 |

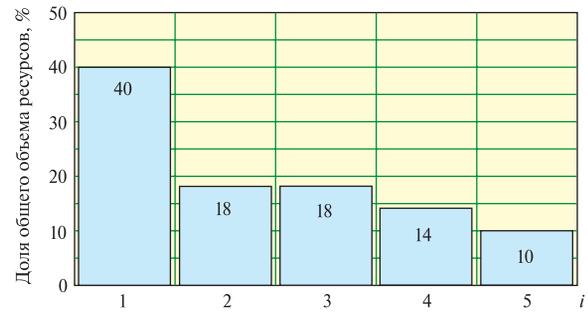


Рис. 2

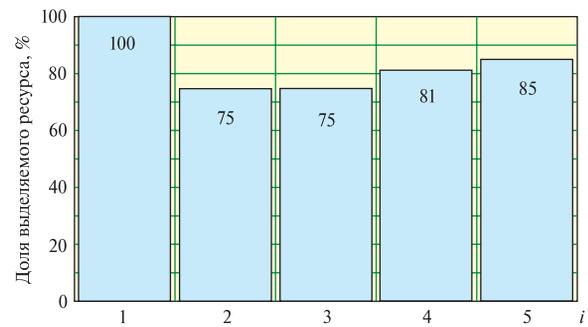


Рис. 3

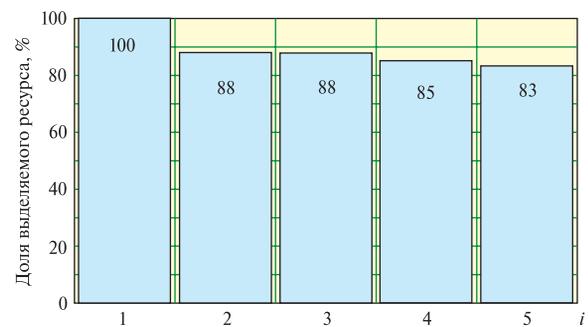


Рис. 4

Модель оптимизации. Модель каждого участка сети может быть описана системой массового обслуживания (СМО) с ожиданием. Таким образом, зависимость качества обслуживания (задержки) от объема сетевых ресурсов (пропускной способности) является нелинейной.

Согласно [14], сеть очередей может рассматриваться как сеть СМО типа М/М/1. Сеть NGN обеспечивает обслуживание трафика с различными приоритетами, поэтому в качестве модели участка сети связи можно рассмотреть СМО с ожиданием и относительными приоритетами. Различные категории пользователей имеют разные приоритеты обслуживания трафика. Показателем качества обслуживания в данном случае может быть время задержки ожидания начала обслуживания [13, 14], определяемое как

$$W_k = \frac{\lambda M(t_S^2)}{2(1 - R_{k-1})(1 - R_k)}, \tag{9}$$

где $R_k = \sum_{i=1}^k A_i$ – интенсивность нагрузки, создаваемой потоком заявок с категориями от 1 до k ; $A_k = \frac{\lambda_k}{\mu_k}$ – интенсивность нагрузки, создаваемой заявками k -й категории; $\lambda = \sum_{i=1}^k \lambda_i$ – суммарная интенсивность потока заявок всех категорий от 1 до k ; $M(t_S^2)$ – второй момент функции распределения времени обслуживания; λ_i – интенсивность трафика поль-

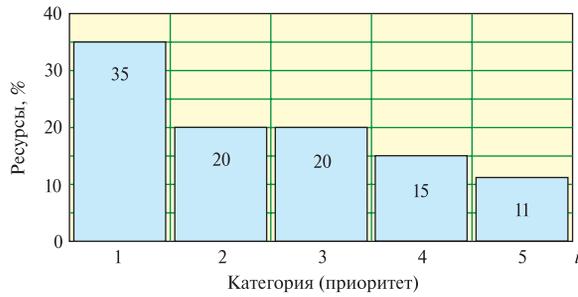


Рис. 5

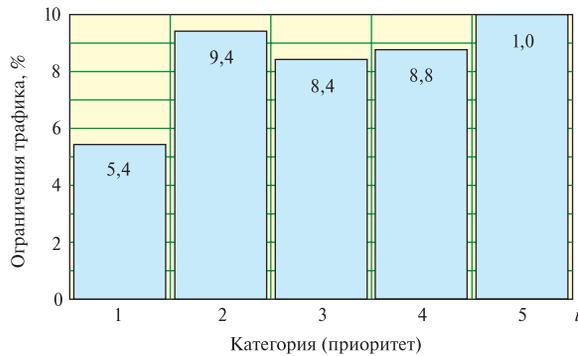


Рис. 6

зователей i -й категории; λ_k – интенсивность трафика пользователей k -й категории; μ_k – интенсивность обслуживания трафика пользователей k -й категории.

Для упрощения модели предположим, что поток пользовательского трафика является простейшим, а время обслуживания имеет экспоненциальное распределение. Тогда $M(t_s^2) = \bar{t}^2$ и соответственно

$$W_k = \frac{\rho \bar{t}}{(1 - R_{k-1})(1 - R_k)}, \tag{10}$$

где $\rho = \lambda \bar{t}$ – интенсивность нагрузки $0 \leq \rho < 1$; \bar{t} – математическое ожидание времени обслуживания.

Если интенсивность обслуживания пользовательского трафика не зависит от его категории, то

$$\mu_1 = \mu_2 = \dots = \mu_k = \mu; \quad A_k = \frac{\lambda_k}{\mu}; \quad R_k = \frac{1}{\mu} \sum_{i=1}^k \lambda_i.$$

Задача распределения сетевых ресурсов возникает при недостатке ресурсов, т.е. при высокой нагрузке: $\rho \rightarrow 1$. Тогда с учетом сделанных допущений можно приближенно записать

$$\tilde{W}_i = \frac{\bar{t}}{\left(1 - \frac{1}{\mu} \sum_{j=1}^{i-1} \lambda_j\right) \left(1 - \frac{1}{\mu} \sum_{j=1}^i \lambda_j\right)}. \tag{11}$$

Для определения функции полезности будем полагать, что доход оператора связи увеличивается с ростом интенсивности пользовательского трафика и уменьшается с увеличением задержки обслуживания. Тогда функция полезности для i -й категории пользователей может быть записана как

$$\varphi_i(\lambda_i) = \frac{\lambda_i}{n_i \lambda_i^{(0)}} \text{ARPU}_i - \frac{\tilde{W}_i}{W_i^{(0)}} p_i, \tag{12}$$

где $\lambda_i^{(0)}$ – интенсивность трафика пользователей i -й категории при достаточном объеме сетевых ресурсов; $W_i^{(0)}$ – средняя величина времени задержки обслуживания пользователей i -й категории.

Используя (12) в выражении (7), методом динамического программирования можно выбрать оптимальное распределение интенсивностей трафика $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_k\}$ пользователей различных категорий, при котором достигается максимум (7). Оптимальный план распределения сетевых ресурсов, полученный на основе исходных данных из приведенного выше примера, представлен на рис. 5.

Оптимальный план распределения ограниченный трафика для различных категорий пользователей приведен на рис. 6.

Выводы.

1. Разработана модель оптимизации распределения сетевых ресурсов для многокатегорийной структуры пользователей сети NGN с ограничениями по качеству обслуживания в соответствии с SLA при использовании классической кусочно-линейной функции полезности.

2. Установлено, что функция полезности для сети очередей не может быть линейной вследствие нелинейной зависимости качества обслуживания (задержки) от объема сетевых ресурсов (пропускной способности) и предложена новая функция полезности.

3. Проведены необходимые расчеты с использованием метода динамического программирования и приведены численные примеры оптимальных планов распределения сетевых ресурсов NGN.

ЛИТЕРАТУРА

1. Гольдштейн Б.С., Соколов Н.А., Яновский Г.Г. Сети связи. – СПб.: БХВ-Петербург, 2010.
2. Росляков А.В. Сети следующего поколения NGN. – М.: Эко-Трендз, 2009.
3. Бакланов И.Г. NGN: принципы построения и организации. – М.: Эко-Трендз, 2008.
4. Кучерявый А.Е., Цуприков А.И. Сети связи следующего поколения. – М.: ЦНИИС, 2006.
5. Гольдштейн Б.С., Гольдштейн А.Б. Softswitch. – СПб.: БХВ-Петербург, 2006.
6. Яновский Г.Г. IP Multimedia Subsystem: принципы, стандарты и архитектура // Вестник связи. – 2006. – № 3.
7. Recommendation Y.2021. IMS for Next Generation Network // ITU-T, 2006.
8. Шварцман В.О. Качество услуг сетей следующего поколения // Электросвязь. – 2006. – № 3.
9. Шнепс-Шнеппе М.А. О соглашениях SLA в условиях NGN и услуг Triple Play // Электросвязь. – 2006. – № 3.
10. ETSI – ETSI EG 202 009-3 V1.2.1. Template for Service Level Agreements (SLA) (2007-01).
11. ITU-T Y.1291. An architectural framework for support of Quality of Service in packet networks. 05/2004.
12. Беллман Р., Дрейфус С. Прикладные задачи динамического программирования. – М.: Наука, 1965.
13. Шнепс-Шнеппе М.А. Системы распределения информации. – М.: Связь, 1979.
14. Клейнрок Л. Теория массового обслуживания. – М.: Машиностроение, 1979.

Получено 26.11.10