# РАСЧЕТ И ОПТИМИЗАЦИЯ СЕТЕЙ С ОЧЕРЕДЯМИ

**Ю. И. Рыжиков (Санкт-Петербург)**

## 1. Постановка задачи

В задачах логистики важнейшую роль играют роль процессы физического и информационного обслуживания грузопотоков [1]. К первым относятся потоки приема транспортных средств, комплектования, пакетирования и погрузки грузов, а также обратные им; ко вторым – соответствующее информационное сопровождение. Реальная логистика всегда сопряжена со *случайными задержками*, которые наносят серьезный экономический ущерб и особенно значимы в условиях флота (прежде всего по погоде, ледовым и гидрологическим условиям, поломкам судов). Управление логистикой *обязано учитывать эти задержки*, для чего необходим адекватный математический аппарат – теория массового обслуживания, она же теория очередей. Первые применения теории очередей к флотским проблемам связываются с именем акад. АН УССР Б.В. Гнеденко (конец 1950-х гг.). Актуальность подобных применений в связи с предполагаемым круглогодичным функционированием Северного морского пути и резким увеличением грузопотока по нему существенно возросла.

Реальные процессы обслуживания практически всегда связаны с прохождением нескольких этапов обслуживания, реализуемых в отдельных узлах *сети*. Проектирование оптимальной сети обслуживания представляет собой сложную комплексную задачу. Ее существенной частью, а также важным элементом оперативного управления сетью, является оптимизация маршрутных матриц для каждого вида неоднородных заявок.

## 2. Сеть обслуживания

Постановку задачи мы опишем сначала для однородного случая. Сеть обслуживания (СеМО) – [2,3,4] - состоит из *рабочих узлов*, занумерованных от 1 до $M$, *источника* (узел «0») и *стока* (узел «$M$+1»). Новая заявка рождается в источнике; с попаданием заявки в сток фиксируется окончание ее пребывания в сети (только для этого он и нужен). Для каждого $j$-го узла сети задаются моменты распределения чистой длительности обслуживания $\{b_{j,l}\}$, $l=\overline{1,L}$, и число каналов $n_j$. Маршрут заявки в сети случаен и определяется неразложимой матрицей передач $R=\{r_{i,j}\}$, $i, j=\overline{0,M+1}$, образованной вероятностями перехода из $i$-го в $j$-й узел.

При анализе сетей обычно предполагается, что распределения длительности обслуживания в узлах определяются только типом заявки и номером узла и не зависят от обслуживания других заявок. Считается также, что вероятности $\{r_{i,j}\}$ перехода заявки из узла $i$ в узел $j$ не зависят от предыстории ее и от состояния узла-преемника $j$.

Точные методы расчета СеМО чрезвычайно громоздки и по условиям известной теоремы ВСМР неприменимы к реальным сетям. Поэтому приближенные методы расчета СеМО опираются на ту или иную форму *декомпозиции* сети. Обычно задачу решают следующим образом:

1. Решением системы уравнений баланса межузловых потоков

$$\lambda_j=\Lambda r_{0,j}+\sum_{i=1}^{M}\lambda_i r_{i,j}, \quad j=\overline{1,M}, \tag{1}$$

где $\Lambda$ – интенсивность потока от источника, вычисляют средние потоки $\{\lambda_i\}$ через узлы.

2. Для всех узлов $j$ проверяют отсутствие перегрузки

$$\lambda_j b_{j,1}/n_j<1, \tag{2}$$

обеспечивающее существование в сети стационарного режима. При его нарушении надлежит скорректировать исходные данные (число каналов в узлах, их быстродействие, маршрутную матрицу).

3. Из сети последовательно выделяют по одному узлы $m = \overline{1, M}$ с присущими им числовыми характеристиками потока и обслуживания. Выделенные узлы рассчитываются как изолированные системы.

4. Определяются показатели распределения времени пребывания заявок для сети в целом.

Расчет отдельных узлов выполняется после фазовой ($H_2$)-аппроксимации распределений длительности обслуживания, позволяющей учесть три момента распределения длительности обслуживания. Методы расчета (итерационный по Такахаси-Таками и матрично-геометрической прогрессии [2,3,4]) весьма трудоемки и сходятся не всегда. Поскольку такой расчет должен выполняться для множества узлов и к тому же в ходе внешних итераций для сети в целом, здесь желательно применение разумных аппроксимаций. Ряд таких аппроксимаций приведен в [4], но без оценок их погрешностей.

Обозначим $\lambda$ – интенсивность входящего в узел простейшего потока, $b$ – среднее время обслуживания, $v$ – коэффициент вариации, $n$ – число каналов в узле. Тогда нужная нам средняя длина очереди может быть выражена через коэффициент загрузки $\rho = \lambda\, b/n$ согласно

$$s = \frac{\rho^2}{1-\rho} \frac{1+v^2}{2n^{1-\rho}}. \tag{3}$$

Среднее число заявок в узле

$$q = s + n\,\rho\,. \tag{4}$$

Качество аппроксимаций (3)-(4) иллюстрирует таблица 1.

Таблица 1. Ожидаемое число заявок в узле

| $n$ | ρ | $v$=0.5 | | $v$=1.0 | | $v$=2.0 | |
|---|---|---|---|---|---|---|---|
| | | точно | ≈ | точно | ≈ | точно | ≈ |
| 1 | 0.7 | 1.721 | 1.721 | 2.333 | 2.333 | 4.783 | 4.783 |
| | 0.8 | 2.800 | 2.800 | 4.000 | 4.000 | 8.800 | 8.800 |
| | 0.9 | 5.963 | 5.963 | 9.000 | 9.000 | 21.150 | 21.150 |
| 3 | 0.7 | 2.841 | 2.834 | 3.249 | 3.275 | 4.841 | 5.037 |
| | 0.8 | 4.049 | 4.005 | 4.989 | 4.969 | 8.689 | 8.822 |
| | 0.9 | 7.337 | 7.236 | 10.054 | 9.957 | 20.841 | 20.843 |
| 5 | 0.7 | 4.079 | 4.130 | 4.382 | 4.508 | 5.543 | 6.020 |
| | 0.8 | 5.429 | 5.450 | 6.216 | 6.319 | 9.285 | 9.798 |
| | 0.9 | 8.851 | 8.810 | 11.362 | 11.396 | 21.200 | 21.740 |
| 10 | 0.7 | 7.351 | 7.512 | 7.517 | 7.819 | 8.089 | 9.047 |
| | 0.8 | 9.079 | 9.262 | 9.637 | 10.019 | 11.410 | 13.048 |
| | 0.9 | 12.878 | 13.021 | 15.019 | 15.434 | 20.098 | 25.085 |

На этапе *агрегации* оцениваются характеристики времени прохождения сети в целом – прежде всего его среднее значение. Для сети в целом справедлива формула Литтла

$$T=Q/\Lambda, \qquad (5)$$

где $Q$ – среднее число заявок в сети и $\Lambda$ – интенсивность входящего внешнего потока. Итак, для получения искомой оценки необходимо найти $Q$.

Оптимизация сети обслуживания должна состоять в улучшении вероятностно-временных характеристик ее работы. Эта задача имеет множество аспектов – в частности, выбор производительности и количества обслуживающих устройств (ледоколов, буксиров, лоцманских судов, причалов, кранов и др.) в узлах. Перечисленные вопросы решаются при проектировании и модернизации сети. *Формальная* оптимизация математической модели сети обслуживания по названным факторам (см., например, [5]) слишком сложна и представляется неконструктивной, поскольку реально здесь приходится производить выбор из *известных конечных наборов значений*. Для заключительных этапов оптимизации особенно актуальна маршрутизация заявок, обеспечивающая минимизацию числа заявок в сети. Эта задача чрезвычайно важна и *в процессе эксплуатации* сетей – при климатических изменениях, смене времен года, изменении гидрологических условий, характера грузопотока, трудоемкостей обработки заявок, выходе из строя части оборудования или персонала и т. п.

## 3. Оптимизация маршрутов в однородной сети

Идею метода целесообразно описать сначала для сети с однородными заявками. Алгоритм оптимизации матрицы передач основан на многошаговом перераспределении потоков заявок из общего предшественника $i$ («отца») от разгружаемого преемника $j$ («сына») к получающему дополнительную нагрузку узлу $k$ («брату»). Перечислим учтенные при этом общие соображения.

1. Максимизация вероятностей передачи заявок из рабочих узлов непосредственно в сток предполагается выполненной на предварительных этапах проектирования сети.
2. Циклические маршруты должны быть исключены. Соответственно далее рассматриваются только ациклические сети, нумерация узлов которых может быть выполнена в соответствии с отношениями предшествования – например, с помощью известного алгоритма Флойда. Это позволяет ограничить объемы перебора в процессе выбора «отца» и «брата».
3. В перераспределении потоков участвуют только дуги, соответствующие *строго положительным* вероятностям перехода.
4. Дополнительная специфика узлов учитывается изменением временных масштабов с помощью коэффициентов $\{c_j\}$.

Метод не учитывает «отдаленных последствий» перераспределения потоков между парой узлов – изменения среднего количества заявок у их преемников. Поэтому он не гарантирует монотонного улучшения итогового показателя.

Алгоритм состоит из следующих этапов:

1. Рассчитать: потоки на входе узлов; коэффициенты загрузки узлов; средние количества $\{q_m\}$ заявок в узлах и их сумму.
2. Подсчитать количество преемников у каждого узла $m$ из множества $\overline{0, M-1}$.
3. Выбрать будущего «сына» – узел $j$ с наибольшим ожидаемым числом заявок в нем.

4. Подобрать ему «отца» – узел $i$ с максимальным потоком в $j$-й узел $\lambda_i r_{i,j}$, имеющий минимум двух преемников.

5. Среди непосредственных преемников «отца» найти «брата» – выбрать узел $k$ по максимально допустимой доле перебрасываемого в него потока $\lambda_i r_{i,j}$. Для узла $m$ эта доля может быть найдена из условия $\rho_m + x \cdot \lambda_i r_{i,j} b_{m,1} / n_m = 1$, откуда следует $x = n_m (1-\rho_m) / b_{m,1} \cdot \lambda_i r_{i,j}$.

6. Индекс «брата» $k$ соответствует максимуму по $m$ отношения $x = n_m (1-\rho_m) / b_{m,1}$ (при выборе максимума произведение $\lambda_i r_{i,j}$ можно не учитывать).

7. Определить максимальную долю $x_2$ потока $\lambda_i r_{i,j}$, перебрасываемую в $k$-й узел, при которой этот узел не будет перегружен. Ограничивая его коэффициент загрузки величиной 0.95, имеем $x_2 = (0.95 - \rho_k) n_k / \lambda_i r_{i,j} b_{k,1}$. При $x_2>1$ назначить $x_2=1$.

8. Найти суммарное число заявок в паре $y_2 = q_j + q_k$. Выполнить аналогичные вычисления для объема переброски $x_1=x_2/2$. Наконец, за $y_0$ принять суммарное число заявок в паре до переброски ($x_0=0$). Зависимость $y(x)$ аппроксимировать параболой $y=Ax^2+Bx+C$, причем $C=y_0$. Для вычисления $A$ и $B$ имеем систему из двух линейных уравнений

$$Ax_1^2 + Bx_1 = y_1 - y_0,$$

$$Ax_2^2 + Bx_2 = y_2 - y_0.$$

Минимум квадратного трехчлена соответствует координате $x^*=-B/2A$. Таким образом,

$$x^* = \frac{(y_1 - y_0)x_2^2 - (y_2 - y_0)x_1^2}{2\left[(y_1 - y_0)x_2 - (y_2 - y_0)x_1\right]}.$$

9. Пересчитать элементы матрицы передач $r_{i,j}$ и $r_{i,k}$.

10. Вернуться к этапу 1.

Работа алгоритма завершается, если изменение среднего числа заявок в сети, вычисляемого на этапе 1, в сравнении с результатом предыдущей итерации оказывается пренебрежимо малым или положительным. В последнем случае в качестве решения принимается результат предыдущего прогона.

## 4. Особенности расчета неоднородной сети

Будем считать, что в сеть из общего источника «0» поступают заявки нескольких типов с моментами распределения обслуживания $\{b_{f,1}, b_{f,2}\}$, $f = \overline{1,F}$. Эти моменты для каждого узла сети корректируются на «местные условия» с помощью масштабных коэффициентов $\{c_m\}$. Каждый тип заявок имеет свою маршрутную матрицу с дополнительным индексом $f$ и, в частности, может обходить некоторые узлы общей («физической») сети. Для таких типов заявок приходится исключать непосещаемые узлы и соответственно корректировать номера переменных, отражая перенумерацию в справочных массивах программы расчета.

На этапе расчета физических узлов производится вычисление суммарных входных потоков и средневзвешенных (моментов распределений обслуживания), по которым рассчитываются среднее число заявок в узле и коэффициент его загрузки.

Выбор «сына» по-прежнему производится по максимуму из $\{q_m\}$. Критическим на каждой итерации типом потока $f^*$ считается создающий в «сыне» максимум по $f$ произведения $\lambda_{f,j} b_{f,1}$. «Отцом» $i$ назначается узел, который по отношению к потоку $f^*$

имеет минимум двух преемников и доставляет по всем узлам $m<j$ максимум произведения $\lambda_{f^*,m} r_{f^*,m,j}$.

Наконец, к «брату» $k$ предъявляются следующие требования:

- элемент матрицы передач $r_{f^*,i,k,}$ должен быть строго положительным (у «сына» и «брата» должен быть общий по $f^*$ «отец»);
- среди допустимых кандидатов $\{m>i\}$, $m \neq j$, он должен иметь минимальное значение $q_m$.

Оптимальная переброска задается четверкой $\{f^*,i,j,k\}$. Ее объем рассчитывается аналогично однородному случаю. Соответственно корректируется интенсивность потока $\lambda_{f^*,j}$ на входе узла $j$. При ее обнулении узел $j$ из маршрута $f^*$-потока исключается, что отражается в справочных таблицах.

## 5. Численный эксперимент

Описанный алгоритм был реализован для условной сети, включающей семь рабочих узлов и четыре типа заявок, для каждого из которых предусматривался обход одного или двух узлов. Коэффициенты вариации времени обслуживания по типам предполагались 2.000, 1.000, 0.815 и 0.577. Ввиду большого объема исходных данных мы приведем только результаты счета (см. таблицу 2).

Таблица 2. Динамика оптимизации неоднородной сети

| iter | Q | ΔQ | Параметры | | | | |
|---|---|---|---|---|---|---|---|
| | | | *f* | *i* | *j* | *k* | *x* |
| 0 | 41.433 | - | 4 | 2 | 5 | 7 | 1.000 |
| 1 | 26.328 | -15.105 | 3 | 2 | 6 | 7 | 1.000 |
| 2 | 24.750 | -1.578 | 1 | 1 | 4 | 5 | 0.670 |
| 3 | 23.409 | -1.341 | – | – | – | – | |

Таблица 2 демонстрирует работу всех ветвей алгоритма, смену критических типов потоков и состава «активных семейств», максимальную и «параболически оптимальную» переброску части потока. Значения $Q$ указаны к началу соответствующей итерации. Таким образом, после двух шагов суммарное число заявок в сети уменьшилось в 1.77 раза. Счет завершился из-за невозможности сформировать «семейство» узлов, уменьшающих целевую функцию.

Напомним, что в соответствии с формулой Литтла относительный выигрыш по среднему времени пребывания заявки в сети в точности равен выигрышу по среднему числу заявок в ней.

## Заключение

В докладе предложен метод минимизации среднего времени пребывания заявки в неоднородной сети обслуживания посредством последовательной минимизации суммарного числа заявок в разгружаемом и нагружаемом узлах. Метод опирается на формулу (3) для средней длины очереди в многоканальных системах и приемлемую по точности в широком диапазоне условий. Алгоритм коррекции маршрутных матриц имеет разумные основания, прост по идее, эффективен и позволяет легко учесть ограничения на допустимость коррекции их отдельных элементов. Потенциальный выигрыш от его применения заметен уже при небольшом числе итераций.

Опыт расчетов систем с очередями [3] показал, что выигрыш в вероятности превышения директивного срока (в значении ДФР) может оказаться *на порядки* больше, чем в среднем времени пребывания. Это обстоятельство делает предложенный метод особенно актуальным для ответственных применений с требованиями быстрого реагирования.

**Литература**


1. Организация и технология перегрузочного процесса: учебное пособие, ч.1. / О.А. Изотов, А.В. Кириченко и др. СПб.: Гос. ун-т морского и речного флота им. С.О. Макарова, 2015. 517 с.
2. **Вишневский В.М.** Теоретические основы проектирования компьютерных сетей. М.: Техносфера, 2003. 512 с.
3. **Рыжиков Ю.И.** Численные методы теории очередей: учебное пособие. СПб.: Издательство «Лань» , 2019. 512 с.
4. **Bolch S., Greiner S., Meer de, H., Trivedi K.S.** Queueing Networks and Markov Chains Modeling and Performance Evaluation with Computer Science Application. N.Y. etc: Wiley and Sons, 1998. 726 pp.
5. **Zadorozhnyi V.N., Kornach M.A.** Semi-Analytical Methods for Complex Optimization of non-Markovian Queueing Networks // Proc. of the 16-th Internat. conf. «Information Technologies and Mathematical Modelling. Queueing Theory and Applications», Kazan. Springer, 2017. P. 382-397.