

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ИССЛЕДОВАНИЯ АЛГОРИТМОВ МАРШРУТИЗАЦИИ В СЕТЯХ ПЕРЕДАЧИ ДАННЫХ

М.П.Березко*, В.М.Вишневский*, Е.В.Левнер**, Е.В.Федотов*

*Институт проблем передачи информации, Российская академия наук, Москва, Россия

**Department of Computer Science, Holon Center for Technological Education, Holon, Israel

Поступила в редколлегию 28.05.2001

1. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Под алгоритмом маршрутизации понимается правило, в соответствии с которым в каждом узле сети передачи данных осуществляется выбор линии связи для передачи блока данных (сообщения или пакета).

В работах [1–6] приводится достаточно подробная классификация алгоритмов маршрутизации. Согласно [2], под **фиксированной (неразветвленной, однопутевой)** маршрутизацией будем понимать такую процедуру выбора маршрутов, при которой для передачи данных от узла-источника узлу-адресату используется единственный маршрут. Если в процедуре выбора маршрутов разрешается использовать более одного пути, то она называется **альтернативной (разветвленной, многопутевой)**.

Очевидно, что в общем случае альтернативная маршрутизация является предпочтительней, чем фиксированная, так как она более полно использует ресурсы сети передачи данных, однако фиксированная маршрутизация намного проще для реализации и в ряде случаев (например, при низкой загрузке сети) при ее использовании качество функционирования сети может оказаться очень близким к варианту с реализацией альтернативной маршрутизации. Частным случаем альтернативной маршрутизации является маршрутизация с ограничением на число исходящих линий **К**, используемых для передачи данных из каждого узла узлу-адресату. Назовем такую маршрутизацию **К-путевой**. Данная маршрутизация достаточно проста для реализации (проще, чем полностью альтернативная) и в более полной мере, чем фиксированная использует ресурсы сети.

На стадии проектирования сети передачи данных и в процессе ее развития очень важной является задача выбора алгоритма маршрутизации. Для этой цели могут быть использованы различные средства:

- измерения и статистический анализ параметров реальной сети (в том случае, если сеть передачи данных уже построена и находится в процессе развития);
- натурное моделирование;
- имитационное моделирование;
- математические оптимизационные модели.

В данной работе описываются математические модели алгоритмов маршрутизации. Применение математического моделирования при анализе алгоритмов маршрутизации в сетях передачи данных представляет значительный интерес по следующим причинам:

- к настоящему времени разработан достаточно полный математический аппарат, позволяющий использовать его как в “чистом виде”, так и осуществлять необходимую модификацию существующих моделей, учитывающую специфику конкретной сети;

– использование математических моделей не требуют значительных ресурсов (вычислительных, временных и т. п.), что позволяет при необходимости проводить многократный анализ в процессе разработки и эксплуатации сети;

– математические модели, в отличие от других вышеперечисленных средств, позволяют делать выводы о тенденциях развития сети передачи данных, что является чрезвычайно важным при построении крупномасштабных сетей передачи данных.

2. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим следующую модель сети передачи данных (СПД). СПД состоит из N узлов коммутации и M линий связи. Предполагается, что:

- 1) все линии связи абсолютно надежны;
- 2) все линии связи помехоустойчивы;
- 3) узлы коммутации имеют бесконечную память;
- 4) время обработки в узлах коммутации отсутствует;
- 5) длины всех сообщений независимы и распределены по показательному закону со средним значением $1/\mu$ [байт];
- 6) трафик, поступающий в сеть, состоит из сообщений, имеющих одинаковый приоритет, и образует пуассоновский поток со средним значением γ_{ij} [сообщений/сек] для сообщений, возникающих в узле i и предназначенных узлу j ; обозначим:

$$\gamma = \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \quad - \quad \text{полный внешний трафик}; \quad (2.1)$$

7) каждая линия связи состоит из единственного дуплексного канала связи с пропускной способностью, равной d_{kl} [байт/сек] ((k, l) – линия связи между узлами k и l); если линия связи между узлами k и l отсутствует, то $d_{kl} = 0$.

Обозначим через $x_{kl}^{(i,j)}$ – долю потока γ_{ij} , проходящую по линии (k, l) :

$$0 \leq x_{kl}^{(i,j)} \leq 1. \quad (2.2)$$

Тогда:

$$\lambda_{kl} = \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)}, \quad (2.3)$$

где λ_{kl} – величина потока в линии (k, l) [сообщений/сек], обусловленная потоком γ_{ij} .

Для переменных $x_{kl}^{(i,j)}$ должно выполняться условие сохранения потока в сети, которое записывается следующим образом:

$$\sum_{k=1}^N x_{kl}^{(i,j)} - \sum_{k=1}^N x_{lk}^{(i,j)} = \begin{cases} -1, & l = i \\ 0, & l \neq i, j \\ 1, & l = j. \end{cases} \quad (2.4)$$

Определим через Z_{ij} – среднее время, затрачиваемое на передачу сообщения, которое возникло в узле i и предназначается узлу j (межконцевая задержка сообщения). Важной характеристикой

качества функционирования сети передачи данных является средняя задержка сообщения в сети – T , которая определяется как взвешенная сумма межконцевых задержек Z_{ij} :

$$T = \frac{1}{\gamma} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} Z_{ij}. \quad (2.5)$$

Применение формулы Литтла к сети очередей приводит к общему, и в то же время чрезвычайно простому результату, впервые полученному Л. Клейнроком [1]:

$$T = \frac{1}{\gamma} \sum_{k=1}^N \sum_{l=1}^N \lambda_{kl} t_{kl}, \quad (2.6)$$

где t_{kl} – среднее время пребывания сообщений в линии (k, l) .

К сожалению, получение аналитического выражения для величины t_{kl} (и, следовательно, для величины T) в общем случае невозможно.

Однако, опираясь на предположения 1) – 7), можно доказать [1, 2], что средняя задержка сообщений в сети T может быть определена аналитически. Действительно, в данном случае сеть очередей сводится к модели, впервые изученной Джексоном [7], в которой каждая линия связи рассматривается как независимая СМО типа М/М/1.

В этом случае, среднее время пребывания сообщений в линии (k, l) , состоящее из времени передачи сообщения – $\frac{1}{\mu d_{kl}}$ и времени ожидания в очереди – W_{kl} определяется по следующей формуле:

$$t_{kl} = \frac{1}{\mu d_{kl}} + W_{kl}, \quad (2.7)$$

где [2, 8]

$$W_{kl} = \frac{1}{\mu d_{kl}} \frac{\lambda_{kl}}{\mu d_{kl} - \lambda_{kl}}$$

или

$$t_{kl} = \frac{1}{\mu d_{kl} - \lambda_{kl}}. \quad (2.8)$$

Обозначим: $f_{kl} = \lambda_{kl}/\mu$ – величина потока в линии (k, l) , выраженная в байтах/сек. Тогда:

$$t_{kl} = \frac{1}{\mu} \frac{1}{d_{kl} - f_{kl}}. \quad (2.9)$$

При подстановке t_{kl} в выражение (2.6) получается выражение для средней задержки сообщений по сети:

$$T = \frac{1}{\gamma} \sum_{k=1}^N \sum_{l=1}^N \frac{f_{kl}}{d_{kl} - f_{kl}} \quad (2.10)$$

Сделанные предположения и обозначения позволяют сформулировать задачу поиска таких значений переменных $x_{kl}^{(i,j)}$, которые обеспечат оптимальное (наименьшее) значение величине T .

Известны:

- 1) топологическая структура СПД;
- 2) матрица входных потоков – $\|\gamma_{ij}\|$;
- 3) пропускные способности линий связи – $\|d_{kl}\|$;
- 4) средняя длина сообщения – $1/\mu$.

Требуется найти:

- 1) переменные $x_{kl}^{(i,j)}$ (см. замечание) и, следовательно, потоки в линиях связи f_{kl} такие, что:

$$T = \frac{1}{\gamma} \sum_{k=1}^N \sum_{l=1}^N \frac{f_{kl}}{d_{kl} - f_{kl}} \rightarrow \min \quad (2.11)$$

при выполнении ограничений:

$$f_{kl} = \frac{1}{\mu} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)}; \quad k, l = 1, 2, \dots, N, \quad (2.12)$$

$$f_{kl} < d_{kl}; \quad k, l = 1, 2, \dots, N, \quad (2.13)$$

$$\sum_{k=1}^N x_{kl}^{(i,j)} - \sum_{k=1}^N x_{lk}^{(i,j)} = \begin{cases} -1, & l = i \\ 0, & l \neq i, j \\ 1, & l = j, \end{cases} \quad (2.14)$$

$$0 \leq x_{kl}^{(i,j)} \leq 1; \quad i, j, k, l = 1, 2, \dots, N. \quad (2.15)$$

Данная задача называется **задачей выбора оптимальных потоков и определения оптимальных маршрутов в сети передачи данных по критерию средней задержки**.

Ограничение (2.15) предполагает, что для передачи сообщений из узла i в узел j может быть использовано более одного маршрута, то есть задача (2.11) – (2.15) описывает **альтернативную** процедуру выбора маршрутов.

Если условие (2.15) заменить на условие

$$x_{kl}^{(i,j)} \in \{0, 1\}; \quad i, j, k, l = 1, 2, \dots, N, \quad (2.15a)$$

то задача (2.11) – (2.14) совместно с условием (2.15a) будет определять **фиксированную** маршрутизацию.

И, наконец, для описания данной задачи для случая **К-путевой** маршрутизации будем использовать следующие дополнительные ограничения.

Введем переменную:

$$\nu_{kl}^{(j)} = \begin{cases} 1, & \text{если } \sum_{i=1}^N x_{kl}^{(i,j)} > 0 \\ 0, & \text{если } \sum_{i=1}^N x_{kl}^{(i,j)} = 0; \end{cases} \quad (2.16)$$

$$j, k, l = 1, 2, \dots, N.$$

Иными словами, переменная $\nu_{kl}^{(j)} = 1$, если линия связи (k, l) используется для передачи потока в узел-адресат j хотя бы от одного узла-источника и равна 0, в противном случае.

Тогда ограничение на число исходящих линий (\mathbf{K}), используемых для передачи данных из каждого узла k узлу-адресату j можно записать в следующем виде:

$$\sum_{l=1}^N \nu_{kl}^{(j)} \leq \mathbf{K}; \quad k, j = 1, 2, \dots, N. \quad (2.17)$$

Таким образом, задача (2.11) – (2.17) описывает **К-путевую** маршрутизацию. Заметим, что если положить величину \mathbf{K} , равной 1, то ограничения (2.16) и (2.17) превратятся в ограничение (2.15а), то есть мы вновь получим постановку задачи для **фиксированной** маршрутизации, что совершенно естественно.

Замечание. Формальным результатом решения задачи выбора оптимальных потоков в сети является множество переменных $x_{kl}^{(i,j)}$; $i, j, k, l = 1, 2, \dots, N$. Зная эти переменные, легко определить величины потоков в линиях связи f_{kl} , множество оптимальных маршрутов для всех пар узлов “источник – адресат” и доли от входящих потоков γ_{ij} , которые нужно передавать по оптимальным маршрутам. Сами переменные $x_{kl}^{(i,j)}$ практического смысла не имеют и многие существующие алгоритмы решения задачи выбора оптимальных потоков, как правило, определяют лишь потоки в линиях связи f_{kl} . Зная значения f_{kl} , по формуле (2.10) можно определить значение минимальной задержки T . Однако, в ряде случаев, необходимо знать, какие именно маршруты приводят к оптимальному распределению потоков.

Более строго, ставится следующая задача: для каждой пары узлов “источник (i) – адресат (j)” необходимо определить множество оптимальных маршрутов $\Pi_{ij} = \{\pi_{ij}^{(r)}\}$, $r = 1, 2, \dots, R_{ij}$ (R_{ij} – количество оптимальных маршрутов из узла i в узел j) и доли потоков $\alpha_{ij}^{(r)}$ от входного потока γ_{ij} , в соответствии с которыми используются маршруты $\pi_{ij}^{(r)}$ ($\sum_{r=1}^N \alpha_{ij}^{(r)} = 1$). Очевидно, что для фиксированной маршрутизации $R_{ij} = 1$ и $\alpha_{ij}^{(1)} = 1$, то есть определяется единственный оптимальный маршрут. Решение данной задачи осуществляется в рамках соответствующих алгоритмов.

3. АЛГОРИТМЫ РЕШЕНИЯ ЗАДАЧИ ВЫБОРА ОПТИМАЛЬНЫХ ПОТОКОВ В СЕТИ

3.1. Альтернативная маршрутизация

В случае альтернативной маршрутизации задача выбора оптимальных маршрутов относится к классу многопродуктовых задач с выпуклой целевой функцией и выпуклым множеством ограничений. Следовательно, существует единственный локальный минимум данной задачи, являющийся глобальным минимумом, для нахождения которого разработано достаточно большое число вычислительных методов [9 – 12].

Наиболее известным методом решения данной задачи является метод отклонения (девиации) потока, предложенный в работе [9]. Данный алгоритм является частным случаем так называемого метода Франка–Вольфе [13] для решения более общих задач нелинейного программирования с выпуклым множеством ограничений. Доказано [9], что метод отклонения потока в пределе уменьшает значение целевой функции до минимума, хотя по мере приближения к оптимуму скорость сходимости существенно замедляется.

Алгоритм отклонения потока опирается на следующие два свойства оптимального решения задачи [9, 12].

Свойство 1. Множество многопродуктовых потоков в линиях связи f является выпуклым многогранником, крайними точками которого являются так называемые “экстремальные” потоки, определяемые в соответствии с принципом кратчайших маршрутов. Данный принцип состоит в том, что в качестве оптимальных маршрутов для каждой пары узлов “источник (i) – адресат (j)” рассматриваются кратчайшие маршруты, вычисляемые при произвольном назначении “весов” линий связи. Каждое такое назначение соответствует некоторому “экстремальному” потоку и наоборот. В работе [9] показано, что любой многопродуктовый поток в линиях связи f может быть представлен, как выпуклая комбинация “экстремальных” потоков.

Свойство 2. Для заданного множества многопродуктовых потоков в линиях связи f определим “вес” линии связи (k, l) как частную производную средней задержки по переменной f_{kl} , то есть $\frac{\partial T}{\partial f_{kl}}$.

Пусть φ – поток по кратчайшим маршрутам, определенным в соответствии с “весами” $\frac{\partial T}{\partial f_{kl}}$. Пусть далее, $\hat{f} = \beta \cdot \varphi + (1 - \beta) \cdot f$ – выпуклая комбинация φ и f , минимизирующая функцию $T(\beta)$. Если $T(\hat{f}) = T(f)$, то f – оптимальный поток.

Справедливость выбора “весов” $\frac{\partial T}{\partial f_{kl}}$ и сходимость алгоритма доказываются нижеследующими рассуждениями.

Пусть: f – оптимальное решение. Это означает, что любое отклонение от f ведет к увеличению целевой функции T , т. е. для любого отклонения \hat{f} относительно f , такого, что \hat{f} – допустимый многопродуктовый поток выполняется неравенство:

$$T(\hat{f}) \geq T(f). \quad (3.1)$$

Если $\hat{f} = \beta \cdot \varphi + (1 - \beta) \cdot f$ – выпуклая комбинация φ и f , где φ – поток по кратчайшим маршрутам, определенным в соответствии с некоторыми “весами” w_{kl} , то

$$\hat{f} = f + \beta \cdot (\varphi - f). \quad (3.2)$$

Рассмотрим разность: $T(\hat{f}) - T(f) = \delta T$. Для $\beta \ll 1$ имеем:

$$T(\hat{f}) - T(f) \approx \sum_{k=1}^N \sum_{l=1}^N (\hat{f}_{kl} - f_{kl}) \frac{\partial T}{\partial f_{kl}} = \beta \sum_{k=1}^N \sum_{l=1}^N (\varphi_{kl} - f_{kl}) \frac{\partial T}{\partial f_{kl}}.$$

Тогда, из неравенства (3.1) следует:

$$\sum_{k=1}^N \sum_{l=1}^N (\varphi_{kl} - f_{kl}) \frac{\partial T}{\partial f_{kl}} \geq 0. \quad (3.3)$$

Или:

$$\sum_{k=1}^N \sum_{l=1}^N \varphi_{kl} \frac{\partial T}{\partial f_{kl}} \geq \sum_{k=1}^N \sum_{l=1}^N f_{kl} \frac{\partial T}{\partial f_{kl}}, \quad (3.4)$$

причем в точке оптимума неравенство (3.4) переходит в равенство. Таким образом, для того чтобы в результате итераций поток f приближался к оптимальному решению, необходимо минимизировать

величину $\sum_{k=1}^N \sum_{l=1}^N \varphi_{kl} \frac{\partial T}{\partial f_{kl}}$, где φ – поток по кратчайшим маршрутам, определенным в соответствии с “весаами” $w_{kl} = \frac{\partial T}{\partial f_{kl}}$.

Ниже предлагается модифицированная версия алгоритма отклонения потока.

Описание алгоритма

Шаг 1. Определить “веса” линий связи w_{kl} и инициализировать потоки в линиях связи f_{kl} :

$$w_{kl} := \begin{cases} \left[\frac{\partial T}{\partial f_{kl}} \right]_{f_{kl}=0} = \frac{1}{d_{kl}}; & k, l = 1, 2, \dots, N : d_{kl} > 0 \\ \infty, & k, l = 1, 2, \dots, N : d_{kl} = 0; \end{cases}$$

$$f_{kl} := 0; \quad k, l = 1, 2, \dots, N.$$

Шаг 2. Используя “веса” линий связи w_{kl} , определить кратчайшие пути π_{ij} между всеми парами узлов “источник – адресат”. Для нахождения кратчайших путей в данном случае наиболее подходящим является алгоритм Флойда [14].

Шаг 3. Распределить потоки по кратчайшим путям:

$$\begin{aligned} &\forall i, j = 1, 2, \dots, N : \\ &\forall (k, l) \in \pi_{ij} : \\ &f_{kl} := f_{kl} + \frac{\gamma_{ij}}{\mu}. \end{aligned}$$

Шаг 4. Вычислить:

$$T_{old} := \frac{1}{\gamma} \sum_{k=1}^N \sum_{l=1}^N \frac{f_{kl}}{d_{kl} - f_{kl}}.$$

Шаг 5. Положить: $\gamma^{(1)} := \gamma$.

Шаг 6. Положить: $\gamma^{(2)} := \min\{\gamma, \frac{\gamma^{(1)}}{\rho_{max}}\}$, где: $\rho_{max} = \max\{\frac{f_{kl}}{d_{kl}}\}$; $\forall (k, l) : d_{kl} > 0$.

Шаг 7. Пересчитать потоки в линиях связи:

$$f_{kl} := f_{kl} \cdot \frac{\gamma^{(2)}}{\gamma^{(1)}}; \quad k, l = 1, 2, \dots, N.$$

Шаг 8. Определить “веса” линий связи w_{kl} и инициализировать потоки по кратчайшим путям φ_{kl} :

$$w_{kl} := \begin{cases} \frac{\partial T}{\partial f_{kl}} = \frac{d_{kl}}{(d_{kl} - f_{kl})^2}; & k, l = 1, 2, \dots, N : d_{kl} > 0 \text{ и } f_{kl} < d_{kl} \\ \infty; & k, l = 1, 2, \dots, N : d_{kl} = 0 \text{ или } f_{kl} \geq d_{kl}; \end{cases}$$

$$\varphi_{kl} := 0; \quad k, l = 1, 2, \dots, N.$$

Шаг 9. Используя “веса” линий связи w_{kl} , определить кратчайшие пути π_{ij} между всеми парами узлов “источник – адресат”.

Шаг 10. Распределить потоки по кратчайшим путям:

$$\begin{aligned} \forall i, j = 1, 2, \dots, N : \\ \forall (k, l) \in \pi_{ij} : \\ \varphi_{kl} := \varphi_{kl} + \gamma_{ij} \cdot \frac{\gamma^{(2)}}{\mu \cdot \gamma}. \end{aligned}$$

Шаг 11. Найти величину $\beta \in [0, 1]$, минимизирующую функцию:

$$T(\beta) = \frac{1}{\gamma^{(2)}} \sum_{k=1}^N \sum_{l=1}^N \frac{\beta \cdot \varphi_{kl} + (1 - \beta) \cdot f_{kl}}{d_{kl} - \beta \cdot \varphi_{kl} - (1 - \beta) f_{kl}}$$

при условии выполнения ограничения (2.13).

Поиск величины β можно осуществить любым из известных методов одномерного поиска, например, методом Фибоначчи. Ограничение (2.13) легко добавить в реализацию метода одномерного поиска: если для некоторого значения β : $\beta \cdot \varphi_{kl} + (1 - \beta) \cdot f_{kl} \geq d_{kl}$, то достаточно положить $T(\beta) = \infty$.

Шаг 12. Выполнить отклонение (девиацию) потока на величину β :

$$f_{kl} = \beta \cdot \varphi_{kl} + (1 - \beta) \cdot f_{kl}; \quad k, l = 1, 2, \dots, N.$$

Шаг 13. Вычислить: $T_{\text{new}} := \frac{1}{\gamma^{(2)}} \sum_{k=1}^N \sum_{l=1}^N \frac{f_{kl}}{d_{kl} - f_{kl}}$.

Шаг 14. Если $|T_{\text{old}} - T_{\text{new}}| \leq \varepsilon$, то STOP:

если $\gamma^{(2)} < \gamma$, то допустимых решений нет;

если $\gamma^{(2)} = \gamma$, то получено оптимальное решение с заданной точностью ε .

Иначе:

1) положить: $T_{\text{old}} := T_{\text{new}}; \quad \gamma^{(1)} := \gamma^{(2)}$;

2) если $\gamma^{(1)} < \gamma$, то перейти к шагу 6; иначе перейти к шагу 8.

По сравнению с исходным описанием [9, 12], алгоритм объединяет в себе шаги построения начального допустимого потока (шаги 1 – 14) и собственно задачи минимизации средней задержки (шаги 8 – 14).

Для определения множества оптимальных маршрутов $\Pi_{ij} = \{\pi_{ij}^{(r)}\}$, $r = 1, 2$, R_{ij} и долей потоков $\alpha_{ij}^{(r)}$ от входного потока γ_{ij} можно использовать модификацию алгоритма, предложенную в работе [15].

В заключение остановимся на теоретической трудоемкости алгоритма. Данная трудоемкость определяется шагами 2 и 9, на которых производится поиск кратчайших путей между всеми парами узлов. Для алгоритма Флойда теоретическая трудоемкость составляет $O(N^3)$, следовательно, трудоемкость алгоритма отклонения потока – $O(N^3)$.

3.2. Фиксированная (однопутевая) маршрутизация

По сравнению с альтернативной маршрутизацией задача поиска единственного оптимального пути является гораздо более сложной. Дело в том, что ограничение (2.15а) является требованием целочисленности переменных $x_{kl}^{(i,j)}$, которое переводит задачу в класс задач целочисленного нелинейного программирования, идентичных задаче “о рюкзаке”. Известно [16], что такие задачи являются NP-полными, то есть для их решения не известны алгоритмы с полиномиальной сложностью. В работах [2, 17, 18] приводятся эвристические алгоритмы, которые позволяют получить приближенное решение задачи с небольшими вычислительными затратами. Одним из основных недостатков эвристических алгоритмов является невозможность оценки погрешности получаемых решений.

В настоящей работе предлагается подход, который позволяет получить точную нижнюю оценку задачи. Данный подход строится на методе релаксации (ослаблении) Лагранжа. Метод состоит в том, что вместо исходной задачи решается двойственная задача, в которой некоторые ограничения включаются в функцию Лагранжа с соответствующими неопределенными множителями (подвергаются лагранжеву ослаблению). Практика показывает [19], что для многих задач целочисленной оптимизации:

- двойственная функция хорошо обусловлена, что позволяет получить удовлетворительную скорость сходимости;

- скачки двойственности (разности между целочисленным оптимумом и оптимумом двойственной задачи) очень малы или могут отсутствовать вовсе.

Таким образом, получение хорошей нижней оценки позволяет:

- оценить качество решений, получаемых эвристическими алгоритмами (и останавливаться на них, если эти решения будут отклоняться от нижней оценки не более чем на заданную величину);

- строить алгоритмы типа “ветвей и границ”, чтобы получать оптимальные решения или решения с заданной точностью.

Использование метода релаксации Лагранжа для решения задачи фиксированной маршрутизации впервые было использовано в работе [16]. Предлагаемый здесь подход использует ослабление меньшего числа ограничений, что повышает точность решения задачи.

Для использования метода релаксации переформулируем исходную задачу. Введем следующее обозначение: $\rho_{kl} = f_{kl}/d_{kl}$ для всех $k, l = 1, 2, \dots, N$. Величина ρ_{kl} называется загрузкой линии связи (k, l) и представляет собой одну из наиболее важных характеристик качества функционирования сети передачи данных. Из ограничения (2.13) следует, что $\rho_{kl} < 1$. Итак:

Требуется найти:

1) Переменные $x_{kl}^{(i,j)}$ и загрузки линий связи ρ_{kl} такие, что:

$$T = \frac{1}{\gamma} \sum_{k=1}^N \sum_{l=1}^N \frac{\rho_{kl}}{1 - \rho_{kl}} \rightarrow \min \tag{3.5}$$

при выполнении ограничений:

$$\rho_{kl} = \frac{1}{\mu \cdot d_{kl}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)}; \quad k, l = 1, 2, \dots, N, \tag{3.6}$$

$$0 \leq \rho_{kl} < 1; \quad k, l = 1, 2, \dots, N, \tag{3.7}$$

$$\sum_{k=1}^N x_{kl}^{(i,j)} - \sum_{k=1}^N x_{lk}^{(i,j)} = \begin{cases} -1, & l = i \\ 0, & l \neq i, j \\ 1, & l = j, \end{cases} \tag{3.8}$$

$$x_{kl}^{(i,j)} \in \{0, 1\}; \quad i, j, k, l = 1, 2, \dots, N. \quad (3.9)$$

Заметим, что целевая функция в (3.5) является строго возрастающей функцией от ρ_{kl} .

Используя этот факт легко доказать, что равенство (3.6) можно заменить на неравенство

$$\frac{1}{\mu \cdot d_{kl}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)} \leq \rho_{kl}; \quad k, l = 1, 2, \dots, N. \quad (3.6a)$$

Теперь исходная задача подготовлена к лагранжеву ослаблению. Избавимся от ограничений (3.6a), умножив их на неопределенные множители Лагранжа σ_{kl} ($\sigma_{kl} \leq 0$; $k, l = 1, 2, \dots, N$) и переходим к функции Лагранжа:

$$L(\sigma) = \frac{1}{\gamma} \sum_{k=1}^N \sum_{l=1}^N \frac{\rho_{kl}}{1 - \rho_{kl}} + \sum_{k=1}^N \sum_{l=1}^N \sigma_{kl} \left\{ \rho_{kl} - \frac{1}{\mu \cdot d_{kl}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)} \right\}.$$

Таким образом, решается следующая задача:

$$L(\sigma) \rightarrow \min \quad (3.10)$$

при выполнении ограничений:

$$0 \leq \rho_{kl} < 1; \quad k, l = 1, 2, \dots, N, \quad (3.11)$$

$$\sum_{k=1}^N x_{kl}^{(i,j)} - \sum_{k=1}^N x_{lk}^{(i,j)} = \begin{cases} -1, & l = i \\ 0, & l \neq i, j \\ 1, & l = j, \end{cases} \quad (3.12)$$

$$x_{kl}^{(i,j)} \in \{0, 1\}; \quad i, j, k, l = 1, 2, \dots, N, \quad (3.13)$$

$$\sigma_{kl} \leq 0; \quad k, l = 1, 2, \dots, N. \quad (3.14)$$

Множество допустимых решений для задачи (3.10) – (3.14) является подмножеством всех допустимых решений задачи (3.5) – (3.9). Условия (3.14) и (3.6a) гарантируют, что для любого допустимого решения задачи (3.5) – (3.9) выражение: $\sigma_{kl} \left\{ \rho_{kl} - \frac{1}{\mu \cdot d_{kl}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)} \right\}$ меньше или равно нулю.

Таким образом, решение задачи (3.10) – (3.14) для произвольных значений переменных σ_{kl} ($\sigma_{kl} \leq 0$) дает возможность определить нижнюю оценку задачи (3.5) – (3.9). Тогда, $\max_{\sigma \leq 0} \{L(\sigma)\}$ также является нижней оценкой задачи (3.5) – (3.9).

Предположим, что значения переменных σ_{kl} известны. Перепишем целевую функцию $L(\sigma)$ в следующем виде:

$$\begin{aligned} L(\sigma) &= \frac{1}{\gamma} \sum_{k=1}^N \sum_{l=1}^N \frac{\rho_{kl}}{1 - \rho_{kl}} + \sum_{k=1}^N \sum_{l=1}^N \sigma_{kl} \left\{ \rho_{kl} - \frac{1}{\mu \cdot d_{kl}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)} \right\} \\ &= \sum_{k=1}^N \sum_{l=1}^N \left\{ \frac{1}{\gamma} \frac{\rho_{kl}}{1 - \rho_{kl}} + \sigma_{kl} \cdot \rho_{kl} \right\} + \sum_{k=1}^N \sum_{l=1}^N \frac{-\sigma_{kl}}{\mu \cdot d_{kl}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)} \\ &= \sum_{k=1}^N \sum_{l=1}^N \left\{ \frac{1}{\gamma} \frac{\rho_{kl}}{1 - \rho_{kl}} + \sigma_{kl} \cdot \rho_{kl} \right\} + \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \sum_{k=1}^N \sum_{l=1}^N \frac{-\sigma_{kl}}{\mu \cdot d_{kl}} \cdot x_{kl}^{(i,j)}. \end{aligned}$$

Данный вид целевой функции позволяет разделить исходную задачу на две независимые подзадачи.

Подзадача 1.

$$\sum_{k=1}^N \sum_{l=1}^N \left\{ \frac{1}{\gamma} \frac{\rho_{kl}}{1 - \rho_{kl}} + \sigma_{kl} \cdot \rho_{kl} \right\} \rightarrow \min \quad (3.15)$$

при выполнении ограничений:

$$0 \leq \rho_{kl} < 1; \quad k, l = 1, 2, \dots, N. \quad (3.16)$$

Подзадача 2.

$$\sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \sum_{k=1}^N \sum_{l=1}^N \frac{-\sigma_{kl}}{\mu \cdot d_{kl}} \cdot x_{kl}^{(i,j)} \rightarrow \min \quad (3.17)$$

при выполнении ограничений:

$$\sum_{k=1}^N x_{kl}^{(i,j)} - \sum_{k=1}^N x_{lk}^{(i,j)} = \begin{cases} -1, & l = i \\ 0, & l \neq i, j \\ 1, & l = j, \end{cases} \quad (3.18)$$

$$x_{kl}^{(i,j)} \in \{0, 1\}; \quad i, j, k, l = 1, 2, \dots, N. \quad (3.19)$$

Подзадача 1 может быть разделена на M независимых подзадач (по одной на каждую линию связи):

$$\frac{1}{\gamma} \frac{\rho_{kl}}{1 - \rho_{kl}} + \sigma_{kl} \cdot \rho_{kl} \rightarrow \min \quad (3.20)$$

при ограничении:

$$0 \leq \rho_{kl} < 1. \quad (3.21)$$

Решение данной подзадачи имеет следующий вид:

$$\rho_{kl} = \begin{cases} 1 - \sqrt{1/(\gamma \cdot \sigma_{kl})}, & \text{если } \sigma_{kl} < -1/\gamma; \\ 0, & \text{в противном случае.} \end{cases} \quad (3.22)$$

Подзадача 2 может быть разделена на N^2 независимых подзадач (по одной на каждую пару “источник (i) – адресат (j)”):

$$\sum_{k=1}^N \sum_{l=1}^N \frac{-\sigma_{kl}}{\mu \cdot d_{kl}} \cdot x_{kl}^{(i,j)} \rightarrow \min \quad (3.23)$$

при выполнении ограничений:

$$\sum_{k=1}^N x_{kl}^{(i,j)} - \sum_{k=1}^N x_{lk}^{(i,j)} = \begin{cases} -1, & l = i \\ 0, & l \neq i, j \\ 1, & l = j, \end{cases} \quad (3.24)$$

$$x_{kl}^{(i,j)} \in \{0, 1\}; \quad i, j, k, l = 1, 2, \dots, N. \quad (3.25)$$

Легко заметить, что задача (3.23)–(3.25) представляет собой задачу выбора кратчайших маршрутов между заданной парой узлов (i, j), если в качестве “весов” ребер рассматривать величины $\frac{-\sigma_{kl}}{\mu \cdot d_{kl}}$. Так как подзадачи (3.23)–(3.25) независимы, то для решения подзадачи 2 можно использовать алгоритм поиска кратчайших путей между всеми парами узлов, например, алгоритм Флойда.

Теперь рассмотрим процедуру поиска множителей Лагранжа σ_{kl} . Функция $L(\sigma)$ является вогнутой, не всюду дифференцируемой функцией от σ [19]. Для данной функции легко вычислить субградиент в точке σ :

$$\eta_{kl} = \rho_{kl} - \frac{1}{\mu \cdot d_{kl}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)}; \quad k, l = 1, 2, \dots, N. \quad (3.26)$$

Тогда, для поиска минимума функции $L(\sigma)$ можно использовать субградиентный алгоритм [20], в соответствии с которым множители Лагранжа σ_{kl} вычисляются по следующей итеративной формуле:

$$\sigma_{kl}^{r+1} = \sigma_{kl}^r - \theta_r \frac{\eta_{kl}^r}{\|\eta_{kl}^r\|}.$$

Существует несколько методов выбора параметров перемещения θ_r , в частности:

- метод с постоянным шагом – $\theta_r = \text{const}$;
- метод сходящегося ряда – $\theta_r = \theta_0(\alpha)^r$;
- метод релаксации – $\theta_r = \delta \frac{L(\sigma_{kl}^r) - \bar{T}}{\|\eta_{kl}^r\|}$ [21], где: \bar{T} – оценка оптимального решения задачи (3.5)–(3.9) и, следовательно, функции $L(\sigma)$; коэффициент релаксации δ удовлетворяет условию: $0 < \delta \leq 2$.

Аналогично [16] остановимся на методе релаксации, который состоит из следующих основных шагов.

Шаг 1. Инициализация:

1) Используя любой приближенный алгоритм, найти верхнюю оценку решения задачи (3.5) – (3.9) – \bar{T} .

2) Выбрать начальные значения множителей Лагранжа – σ_{kl}^0 ; $k, l = 1, 2, \dots, N$.

3) Установить:

$r := 0$ (r – номер итерации);

$p := 0$ (p – счетчик “безрезультатных” итераций);

$\delta := 2$ (коэффициент релаксации);

$L^{opt} := 0$ (текущее значение оптимального решения).

Шаг 2. Решение задачи Лагранжа:

Используя значения σ_{kl}^r решить подзадачи 1 и 2. В результате решения будут получены: $L(\sigma_{kl}^r)$, ρ_{kl} и $x_{kl}^{(i,j)}$; $i, j, k, l = 1, 2, \dots, N$.

Шаг 3. Оценка полученного решения и проверка условий прекращения поиска:

1) Если $L(\sigma_{kl}^r) > L^{opt}$, то положить:

a) $L^{opt} := L(\sigma_{kl}^r)$;

b) $\sigma_{kl}^* := \sigma_{kl}^r$ (множители Лагранжа, соответствующие текущему оптимальному решению); $k, l = 1, 2, \dots, N$;

c) $p := 0$.

2) Если $x_{kl}^{(i,j)}$ являются допустимыми решениями задачи (3.5) – (3.9), то вычислить среднее значение задержки сообщения T , соответствующее переменным $x_{kl}^{(i,j)}$ и, если, $T < \bar{T}$, то запомнить данную величину в качестве оптимального решения: $\bar{T} := T$.

3) Положить $p := p + 1$. Если $p \geq$ заданного значения p_{max} , то:

a) $\delta := \delta/2$;

b) $\sigma_{kl}^r := \sigma_{kl}^*$; $k, l = 1, 2, \dots, N$;

c) $p := 0$;

d) $r := r + 1$;

e) перейти к шагу 2.

4) Если $r \geq$ заданного значения r_{max} , или $\delta \leq \delta_{max}$, или $|T - L^{opt}|/L^{opt} \leq$ заданной погрешности вычислений ε , то STOP: найдено оптимальное решение.

Шаг 4. Вычисление множителей Лагранжа:

1) Вычисляется новое значение субградиента:

$$\eta_{kl} = \rho_{kl} - \frac{1}{\mu \cdot d_{kl}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \cdot x_{kl}^{(i,j)}; \quad k, l = 1, 2, \dots, N.$$

2) Вычисляется параметр перемещения θ_r :

$$\theta_r = \delta \frac{L(\sigma_{kl}^r) - \bar{T}}{\|\eta_{kl}^r\|}.$$

3) Вычисляются новые множители Лагранжа:

$$\sigma_{kl}^{r+1} = \sigma_{kl}^r - \theta_r \frac{\eta_{kl}^r}{\|\eta_{kl}^r\|}.$$

- 4) $r := r + 1$.
- 5) Переход к шагу 2.

Как уже отмечалось выше, получение точной нижней оценки для задачи поиска единственного оптимального пути представляет собой очень важную задачу, однако, в общем случае решение данной задачи может не удовлетворять требуемым ограничениям. В тех случаях, когда в результате решения задачи необходимо получить оптимальные пути между каждой парой узлов “источник – адресат”. Для этой цели необходимо использовать приближенные методы, один из которых описывается ниже. Заметим также, что решение, полученное с помощью данного алгоритма, может быть использовано на шаге 1 предыдущего алгоритма.

Шаг 1. Инициализация:

1) Упорядочить множество пар “источник (i) – адресат (j)” по убыванию величин γ_{ij} . Полученное множество пар обозначим через $\Omega^{(1)}$.

2) Положить $f_{kl} := 0$; $k, l = 1, 2, \dots, N$.

Шаг 2. Поиск допустимого решения:

1) Для \forall пары “источник (i) – адресат (j)” $\in \Omega^{(1)}$ выполнить следующие действия:

а) найти “наименее загруженный” маршрут π_{ij}^0 (под загрузкой маршрута π_{ij} понимается величина $\max_{(k,l) \in \pi_{ij}} \{\rho_{kl}\}$); если таких маршрутов несколько, выбрать среди них маршрут с наименьшим числом промежуточных узлов;

б) распределить потоки по выбранному маршруту:

$$\forall (k, l) \in \pi_{ij}^0 : f_{kl} := f_{kl} + \frac{\gamma_{ij}}{\mu},$$

если $\exists (k, l) : f_{kl} \geq d_{kl}$ то STOP (допустимого решения не существует).

2) Вычислить задержки Z_{ij} между каждой парой узлов “источник (i) – адресат (j)”.

Шаг 3. Поиск “оптимального” решения:

1) Упорядочить множество пар “источник (i) – адресат (j)” по убыванию величин $\gamma_{ij} Z_{ij}$. Полученное множество пар обозначим через $\Omega^{(2)}$.

2) Для \forall пары “источник (i) – адресат (j)” $\in \Omega^{(2)}$ выполнить следующие действия:

а) найти маршрут π_{ij}^{opt} , такой, что отклонение всего потока γ_{ij} на этот маршрут приведет к максимальному уменьшению величины Z_{ij} (алгоритм определения такого маршрута будет описан ниже);

б) распределить потоки по выбранному маршруту:

$$\forall (k, l) \in \pi_{ij}^{opt} : f_{kl} := f_{kl} + \frac{\gamma_{ij}}{\mu}.$$

3) Вычислить: $T := \frac{1}{\gamma} \sum_{k=1}^N \sum_{l=1}^N \frac{f_{kl}}{d_{kl} - f_{kl}}$.

Остановимся подробнее на следующих шагах приближенного алгоритма: определение “наименее загруженного” маршрута (шаг 2.1а) и определение “оптимального” маршрута (шаг 3.2а).

“Наименее загруженный” маршрут можно найти с помощью модификации алгоритма Дейкстры [22].

Для определения “оптимального” маршрута можно использовать следующий простой алгоритм (напомним, что определение такого маршрута выполняется для заданной пары узлов (i, j)).

Шаг 1. Удалить поток γ_{ij} из “наименее загруженного” маршрута π_{ij}^0 (предполагается, что все маршруты π_{ij}^0 были сохранены):

$$\forall (k, l) \in \pi_{ij}^0 : f_{kl} := f_{kl} + \frac{\gamma_{ij}}{\mu}.$$

Шаг 2. Добавить поток γ_{ij} во все линии связи (k, l) :

$$f_{kl} := f_{kl} + \frac{\gamma_{ij}}{\mu}; \quad k, l = 1, 2, \dots, N : d_{kl} > 0.$$

Шаг 3. Вычислить “веса” линий связи (k, l) :

$$\text{если } f_{kl} < d_{kl}, \text{ положить: } w_{kl} := \frac{1}{\mu} \frac{1}{d_{kl} - f_{kl}},$$

иначе: $w_{kl} := \infty$.

Шаг 4. Используя “веса” w_{kl} определить кратчайший путь между парой узлов $(i, j) - \pi_{ij}^{opt}$. Для этой цели вновь можно использовать алгоритм Дейкстры.

Шаг 5. Удалить поток γ_{ij} из всех линии связи (k, l) :

$$f_{kl} := f_{kl} + \frac{\gamma_{ij}}{\mu}; \quad k, l = 1, 2, \dots, N : d_{kl} > 0.$$

Шаг 6. Конец работы алгоритма.

Теоретическая сложность приближенного алгоритма определяется шагами 2.1а и 3.2а и составляет $O(N^4)$.

3.3. К-путевая маршрутизация

При решении данной задачи будем предполагать, что ограничение на число исходящих линий – K больше 1, т. к. случай $K = 1$ рассматривался в разделе 3.2. При $K > 1$ задача (2.11) – (2.17) избавляется от целочисленности переменных $x_{kl}^{(i,j)}$, что делает ее более простой, чем задача фиксированной маршрутизации. Однако наличие дополнительного ограничения несколько усложняет данную задачу по сравнению со случаем альтернативной маршрутизации.

Для решения задачи (2.11) – (2.17) предлагается алгоритм, являющийся модификацией алгоритма отклонения потока. Модификация заключается в том, что отклонение потока выполняется не для всех пар “источник-адресат” одновременно, а для каждой пары отдельно. Похожие идеи использованы в работе [23] при описании алгоритма решения задачи альтернативной маршрутизации.

Описание алгоритма

Шаг 1. Определить “веса” линий связи w_{kl} и инициализировать потоки в линиях связи f_{kl} :

$$w_{kl} := \begin{cases} \left[\frac{\partial \Gamma}{\partial f_{kl}} \right]_{f_{kl}=0} = \frac{1}{d_{kl}}; & k, l = 1, 2, \dots, N : d_{kl} > 0 \\ \infty, & k, l = 1, 2, \dots, N : d_{kl} = 0; \end{cases}$$

$$f_{kl} := 0; \quad k, l = 1, 2, \dots, N.$$

Шаг 2. Положить:

$$\Pi_{ij} := \emptyset \quad (\Pi_{ij} - \text{множество оптимальных маршрутов между парой узлов } (i, j)), \quad i, j = 1, 2, \dots, N;$$

$$\nu_{kl}^{(j)} := 0; \quad k, l, j = 1, 2, \dots, N.$$

Шаг 3. Используя “веса” линий связи w_{kl} , определить кратчайшие пути π_{ij}^0 между всеми парами узлов “источник – адресат”.

Включить кратчайшие пути π_{ij}^0 в множества Π_{ij} :

$$\pi_{ij}^0 \Rightarrow \Pi_{ij}; \quad i, j = 1, 2, \dots, N.$$

Шаг 4. Распределить потоки по кратчайшим путям: $\forall i, j = 1, 2, \dots, N : \forall (k, l) \in \pi_{ij}^0$;

$$1) f_{kl} := f_{kl} + \frac{\gamma_{ij}}{\mu};$$

2) если $\nu_{kl}^{(j)} = 0$, то $\nu_{kl}^{(j)} = 1$.

$$\text{Шаг 5. Вычислить: } T_{\text{old}} := \frac{1}{\gamma} \sum_{k=1}^N \sum_{l=1}^N \frac{f_{kl}}{d_{kl} - f_{kl}}.$$

Шаг 6. Положить: $\gamma^{(1)} := \gamma$.

Шаг 7. Положить: $\gamma^{(2)} := \min\{\gamma, \frac{\gamma^{(1)}}{\rho_{\max}}\}$, где: $\rho_{\max} = \max\{\frac{f_{kl}}{d_{kl}}\}$; $\forall (k, l) : d_{kl} > 0$.

Шаг 8. Пересчитать потоки в линиях связи:

$$f_{kl} := f_{kl} \cdot \frac{\gamma^{(2)}}{\gamma^{(1)}}; \quad k, l = 1, 2, \dots, N.$$

Шаг 9. Вычислить задержки Z_{ij} между каждой парой узлов “источник (i) – адресат (j)”.

Шаг 10. Отсортировать множество пар узлов (i, j) в порядке убывания величин $\gamma_{ij} Z_{ij}$. Полученное множество пар обозначим через Ω .

Шаг 11. Выбрать очередную пару узлов $(i_0, j_0) \in \Omega$. Если все пары рассмотрены, то STOP:

если $\gamma^{(2)} < \gamma$, то допустимых решений нет;

если $\gamma^{(2)} = \gamma$, то получено оптимальное решение с заданной точностью ε .

Шаг 12. Для выбранной пары узлов (i_0, j_0) инициализировать потоки по кратчайшим путям φ_{kl} :

$$1) \varphi_{kl} := f_{kl}; \quad k, l = 1, 2, \dots, N;$$

$$2) \forall (k, l) \in \pi_{i_0 j_0} : \pi_{i_0 j_0} \in \Pi_{i_0 j_0} : \varphi_{kl} := \varphi_{kl} - \gamma_{i_0 j_0} \cdot \frac{\gamma^{(2)}}{\mu \cdot \gamma}.$$

Шаг 13. Определить “веса” линий связи w_{kl} :

$$w_{kl} := \begin{cases} \frac{\partial T}{\partial f_{kl}} = \frac{d_{kl}}{(d_{kl} - f_{kl})^2}; & k, l = 1, 2, \dots, N : d_{kl} > 0 \text{ и } f_{kl} < d_{kl} \\ \infty; & k, l = 1, 2, \dots, N : d_{kl} = 0 \text{ или } f_{kl} \geq d_{kl}. \end{cases}$$

Шаг 14. Используя “веса” линий связи w_{kl} , определить кратчайший путь $\pi_{i_0 j_0}^0$ для пары узлов (i_0, j_0) .

Шаг 15. Проверить ограничение (2.17):

$$1) \forall (k, l) \in \pi_{i_0 j_0}^0 :$$

$$a) \eta_{kl}^{(j_0)} := \nu_{kl}^{(j_0)};$$

b) если $\eta_{kl}^{(j_0)} = 0$, то $\eta_{kl}^{(j_0)} = 1$.

2) $\forall k : (k, l) \in \pi_{i_0 j_0}^0$ определить:

$$a) \kappa = \sum_{l=1}^N \eta_{kl}^{(j_0)};$$

b) если $\kappa > K$ (условие (2.17) нарушено), то перейти к шагу 11.

Шаг 16. Распределить поток $\gamma_{i_0 j_0}$ по кратчайшему пути $\pi_{i_0 j_0}^0$:

$\forall (k, l) \in \pi_{i_0 j_0}^0$:

$$\varphi_{kl} := \varphi_{kl} + \gamma_{i_0 j_0} \cdot \frac{\gamma^{(2)}}{\mu \cdot \gamma}.$$

Шаг 17. Найти величину $\beta \in [0, 1]$, минимизирующую функцию:

$$T(\beta) = \frac{1}{\gamma^{(2)}} \sum_{k=1}^N \sum_{l=1}^N \frac{\beta \cdot \varphi_{kl} + (1 - \beta) \cdot f_{kl}}{d_{kl} - \beta \cdot \varphi_{kl} - (1 - \beta) \cdot f_{kl}}$$

при условии выполнения ограничения (2.13).

Шаг 18. Выполнить отклонение потока на величину β :

$$f_{kl} = \beta \cdot \varphi_{kl} + (1 - \beta) \cdot f_{kl}; \quad k, l = 1, 2, \dots, N.$$

Шаг 19. Вычислить: $T_{\text{new}} := \frac{1}{\gamma^{(2)}} \sum_{k=1}^N \sum_{l=1}^N \frac{f_{kl}}{d_{kl} - f_{kl}}$.

Шаг 20. Если $|T_{\text{old}} - T_{\text{new}}| \leq \varepsilon$, то перейти к шагу 11 (значение целевой функции улучшить не удалось).

Иначе:

1) Включить кратчайший путь $\pi_{i_0 j_0}^0$ в множество $\Pi_{i_0 j_0}$:

$$\pi_{i_0 j_0}^0 \Rightarrow \Pi_{i_0 j_0}.$$

2) $\forall (k, l) \in \pi_{i_0 j_0}^0 : \nu_{kl}^{(j_0)} := \eta_{kl}^{(j_0)}$.

3) Положить: $T_{\text{old}} := T_{\text{new}}; \gamma^{(1)} := \gamma^{(2)}$.

4) если $\gamma^{(1)} < \gamma$, то перейти к шагу 7; иначе перейти к шагу 9.

4. ПРИМЕРЫ АНАЛИЗА АЛГОРИТМОВ МАРШРУТИЗАЦИИ В СЕТЯХ ПЕРЕДАЧИ ДАННЫХ

Предложенные выше математические модели были использованы для анализа двух ведомственных сетей передачи данных, являющихся основой систем продажи и бронирования железнодорожных и авиационных билетов – “Экспресс” и “Сирена”.

4.1. Анализ различных вариантов алгоритмов маршрутизации для СПД “Экспресс”

Сеть передачи данных “Экспресс” состоит из 30 узлов коммутации и 76 линий связи. В таблице 4.1 приведен перечень городов – мест размещения узлов коммутации сети, таблица 4.2 описывает топологическую структуру СПД и характеристики линий связи между узлами коммутации.

Таблица 4.1. Перечень мест размещения узлов коммутации СПД “Экспресс”.

№ п/п	Города	№ п/п	Города
1	Москва	16	Иркутск
2	Санкт-Петербург	17	Рига
3	Киев	18	Львов
4	Кишинев	19	Ярославль
5	Екатеринбург	20	Саратов
6	Харьков	21	Донецк
7	Одесса	22	Красноярск
8	Самара	23	Алма-Ата
9	Симферополь	24	Актюбинск
10	Ростов	25	Акмола
11	Минск	26	Воронеж
12	Н.Новгород	27	Калининград
13	Ташкент	28	Хабаровск
14	Челябинск	29	Киев-2
15	Новосибирск	30	Чита

Сообщения, поступающие в СПД “Экспресс”, подразделяются на два класса: диалоговые сообщения и архивные сообщения, причем диалоговые сообщения имеют относительный приоритет (приоритет без прерывания обслуживания) по отношению к архивным. Средняя длина диалоговых сообщений составляет 167 байтов, средняя длина архивных сообщений – 1524 байта. В результате измерений на реальной сети были определены матрицы входных потоков для каждого класса.

Первоначально в СПД “Экспресс” использовался алгоритм с однопутевой стратегией маршрутизации. В процессе развития системы “Экспресс” требования к надежности и пропускной способности сети передачи данных резко повысились, что было связано как со значительным увеличением узлов и ребер сети (от 3-х узловой сети в 1988 г. до сети из 30 узлов в настоящее время), так и с внедрением новых технологий организации продажи билетов и управления пассажирскими перевозками. Одним из недостатков используемого алгоритма являлось отсутствие расщепления трафика между несколькими маршрутами и, как следствие, неравномерность распределения трафика между линиями связи. В связи с этим возникла необходимость разработки алгоритма, позволяющего более эффективно использовать ресурсы сети передачи данных и тем самым увеличить ее пропускную способность.

Одним из вариантов такого алгоритма является алгоритм **К**-путевой (**К**=2) маршрутизации, в котором устанавливаются два направления передачи к одному и тому же узлу, используемые попеременно.

Для того, чтобы оценить целесообразность разработки и внедрения алгоритма 2-х путевой маршрутизации было организовано сравнение следующих методов маршрутизации:

- однопутевой;
- **К**-путевой (**К**=2);
- альтернативной.

Модель альтернативной маршрутизации использовалась для поиска оптимального решения, позволяющего определить “идеальные” характеристики СПД.

Сравнение методов маршрутизации проводилось двумя способами.

Во-первых, были проведены расчеты на реальных исходных данных.

Результаты расчетов содержат следующую информацию:

- 1) среднее время реакции диалоговых сообщений по сети:
 - для альтернативной маршрутизации – 8.45 сек;

Таблица 4.2. Топологическая структура и характеристики каналов связи СПД “Экспресс”.

№ п/п	Линия связи	Пропускная способность линии связи (бит/сек)	Число каналов связи	№ п/п	Линия связи	Пропускная способность линии связи (бит/сек)	Число каналов связи
1	Москва – Санкт-Петербург	9600	4	39	Харьков – Донецк	1200	1
2	– Киев	2400	1	40	– Воронеж	1200	1
3	– Кишинев	2400	1	41	– Киев-2	2400	1
4	– Екатеринбург	9600	2	42	Одесса – Симферополь	1200	1
5	– Харьков	2400	2	43	– Львов	1200	1
6	– Самара	9600	2	44	– Киев-2	1200	1
7	– Симферополь	9600	1	45	Самара – Н.Новгород	8400	2
8	– Ростов	8400	3	46	– Челябинск	7200	2
9	– Минск	9600	2	47	– Саратов	1200	4
10	– Н.Новгород	9600	2	48	– Воронеж	2400	1
11	– Челябинск	4800	1	49	Симферополь – Донецк	1200	1
12	– Иркутск	9600	1	50	– Киев-2	9600	1
13	– Ярославль	2400	2	51	Ростов – Н.Новгород	4800	1
14	– Донецк	2400	1	52	– Саратов	1200	2
15	– Воронеж	9600	2	53	– Донецк	2400	1
16	– Калининград	9600	1	54	– Воронеж	9600	2
17	– Хабаровск	2400	1	55	Минск – Рига	2400	2
18	– Киев-2	9600	1	56	– Калининград	9600	1
19	Санкт-Петербург – Киев	2400	1	57	Н.Новгород – Ярославль	1200	2
20	– Екатеринбург	9600	1	58	Ташкент – Алма-Ата	1200	1
21	– Н.Новгород	9600	1	59	– Актобинск	1200	1
22	– Рига	2400	1	60	Челябинск – Новосибирск	2400	1
23	– Ярославль	1200	2	61	– Алма-Ата	4800	1
24	– Киев-2	7200	4	62	– Актобинск	2400	1
25	Киев – Кишинев	1200	1	63	– Акмола	2400	2
26	– Одесса	1200	1	64	Новосибирск – Иркутск	1200	1
27	– Минск	2400	1	65	– Красноярск	2400	1
28	– Львов	2400	1	66	– Алма-Ата	1200	1
29	– Киев-2	7200	4	67	Иркутск – Красноярск	2400	1
30	Кишинев – Одесса	1200	1	68	– Хабаровск	1200	1
31	Екатеринбург – Н.Новгород	7200	2	69	– Чита	9600	2
32	– Челябинск	19200	2	70	Рига – Калининград	1200	2
33	– Новосибирск	1200	1	71	Львов – Киев-2	2400	1
34	– Красноярск	2400	1	72	Саратов – Актобинск	1200	1
35	– Акмола	1200	1	73	– Воронеж	1200	1
36	Харьков – Одесса	1200	1	74	Алма-Ата – Акмола	1200	1
37	– Симферополь	1200	1	75	Хабаровск – Чита	1200	1
38	– Ростов	1200	1				

- для 2-х путевой маршрутизации – 8.49 сек;
- для фиксированной (однопутевой) маршрутизации – 8.50 сек;
- 2) максимальное время реакции среди всех пар узлов коммутации:
 - для альтернативной маршрутизации – 39.73 сек;
 - для 2-х путевой маршрутизации – 41.88 сек;
 - для фиксированной (однопутевой) маршрутизации – 42.59 сек;
- 3) средняя загрузка в линиях связи:
 - для альтернативной маршрутизации – 0.12;
 - для 2-х путевой маршрутизации – 0.12;
 - для фиксированной (однопутевой) маршрутизации – 0.12;
- 4) максимальная загрузка в линиях связи:
 - для альтернативной маршрутизации – 0.51;

- для 2-х путевой маршрутизации – 0.51;
- для фиксированной (однопутевой) маршрутизации – 0.54.

Анализ полученных результатов показал преимущество 2-х путевой маршрутизации по сравнению с фиксированной, и альтернативной маршрутизации по сравнению с остальными. Однако, значения основных характеристик отличаются незначительно, что объясняется низкой нагрузкой на сеть (средняя загрузка в линиях связи составляет 0.12).

Для более полного исследования были проведены численные эксперименты, при которых элементы матриц входных потоков одновременно умножались на одну и ту же величину – нагрузку на сеть (начальное значение этой величины принимается равным 1). В результате экспериментов были построены функциональные зависимости средней задержки сообщений по сети от величины нагрузки. Графики данных зависимостей приведены на рис. 4.1. Результаты экспериментов показывают, что перегрузка сети наступает при увеличении нагрузки на сеть:

- для фиксированной маршрутизации - в 1.7 раза;
- для 2-х путевой маршрутизации – в 2.6 раза;
- для альтернативной маршрутизации – в 3.1 раза.

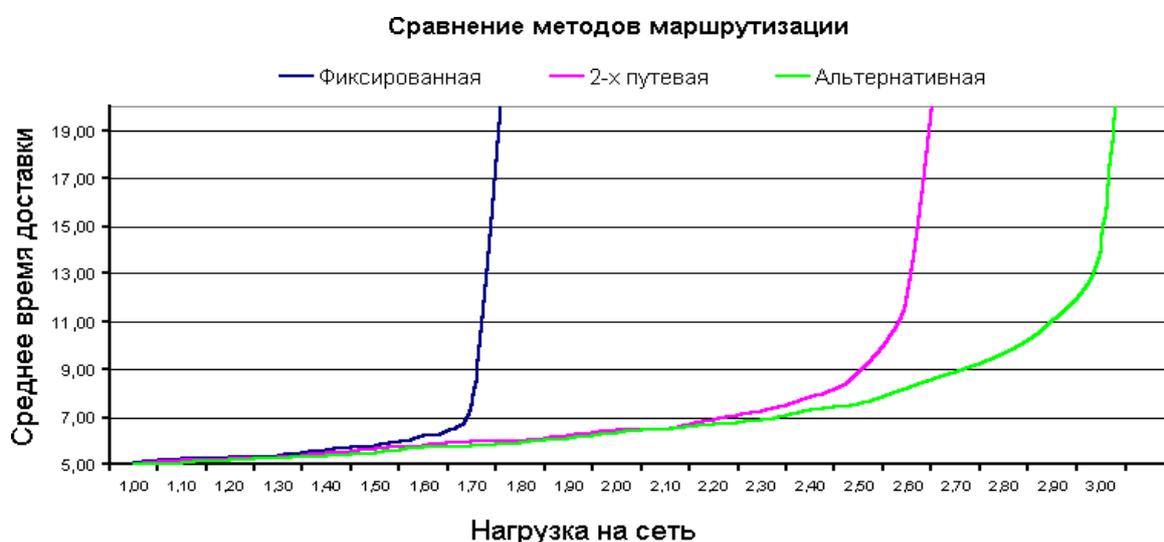


Рис. 4.1.

Таким образом, 2-х путевая маршрутизация допускает значительно большую нагрузку на сеть, чем фиксированная (однопутевая), и в тоже время, характеристики СПД для 2-х путевой маршрутизации очень близки к случаю альтернативной маршрутизации (фактически, при увеличении нагрузки до 2.5 раз), то есть 2-х путевая маршрутизация достаточно близка к идеальной (оптимальной) маршрутизации.

Проведенный анализ с использованием математических моделей, а также результаты измерений на реальной сети, позволили сделать вывод о целесообразности реализации алгоритма 2-х путевой маршрутизации.

4.2. Анализ развития СПД “Сирена”

Сеть передачи данных “Сирена” является основой для распределенного взаимодействия центров систем “Сирена-2.3”, “Сирена-2000” и “Сирена-3”, осуществляющих продажу и бронирование авиабилетов.

В настоящее время сеть передачи данных “Сирена” состоит из 42 магистральных и 34 региональных узлов, размещенных на территории России и других государств СНГ. В таблице 4.3 приведен перечень городов – мест размещения узлов коммутации СПД, в таблице 4.4 – топологическая структура СПД и характеристики линий связи между узлами коммутации.

Таблица 4.3. Перечень мест размещения узлов коммутации СПД “Сирена”.

№ п/п	Города	№ п/п	Города
1	Москва (6 УК)	17	Хабаровск (2 УК)
2	Санкт-Петербург	18	Краснодар
3	Киев	19	Сочи
4	Екатеринбург	20	Тюмень (3 УК)
5	Самара	21	Волгоград
6	Симферополь	22	Казань
7	Ростов (2 УК)	23	Мурманск
8	Минск	24	Братск
9	Ташкент	25	Якутск
10	Челябинск	26	Владивосток
11	Новосибирск (2 УК)	27	Бишкек
12	Иркутск	28	Минеральные Воды
13	Львов	29	Уфа
14	Донецк	30	Архангельск
15	Красноярск (2 УК)	31	Петропавловск – К.
16	Алма-Ата		

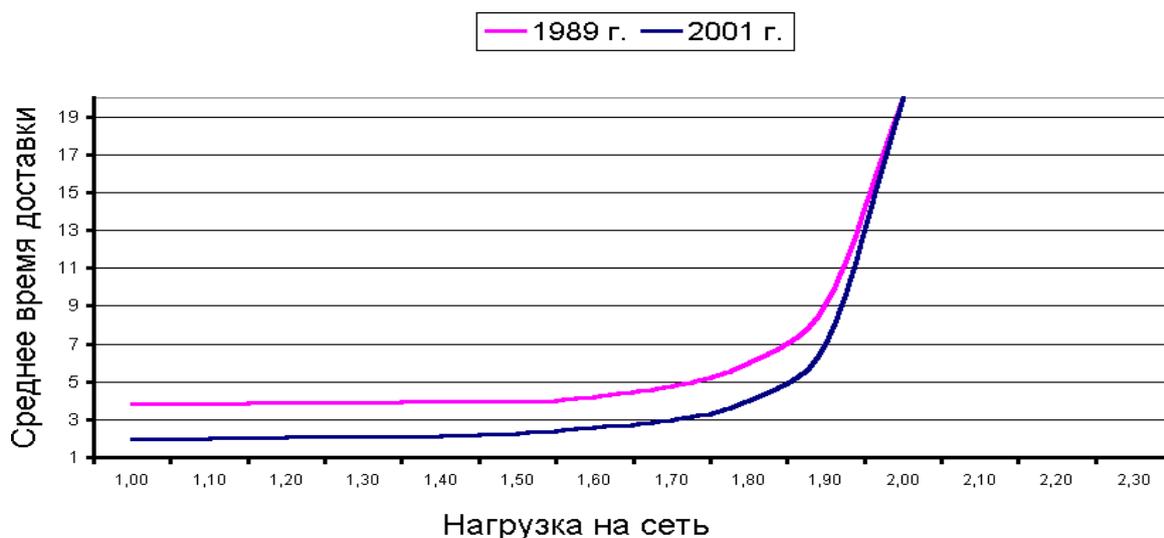


Рис. 4.2.

В течение длительного периода своего развития (с 1981 г. по настоящее время) сеть “Сирена” претерпела значительные изменения. Для оценки тенденций развития СПД “Сирена” были проведены численные эксперименты, в которых осуществлялось сравнение среднего времени доставки по сети для текущего состояния СПД и состояния на 1989 год. В качестве математической модели СПД была использована оптимизационная модель, в которой решалась задача выбора оптимальных маршрутов передачи пакетов по сети и распределения потоков с альтернативной (разветвленной) стратегией маршрутизации.

Таблица 4.4. Топологическая структура и характеристики каналов связи СПД “Сирена”.

№ п/п	Линия связи	Пропускная способность линии связи (бит/сек)	Число каналов связи	№ п/п	Линия связи	Пропускная способность линии связи (бит/сек)	Число каналов связи
1	Москва – Санкт-Петербург	32К + 9600	2	26	Санкт-Петербург – Ростов	9600	1
2	– Киев	9600	2	27	– Новосибирск	9600	1
3	– Екатеринбург	19200+ 9600	2	28	– Мурманск	9600	1
4	– Самара	19200	1	29	– Архангельск	9600	1
5	– Симферополь	9600	2	30	Киев – Львов	9600	1
6	– Ростов	64К+9600	2	31	– Донецк	9600	1
7	– Минск	19200+9600	2	32	Екатеринбург – Самара	9600	1
8	– Ташкент	9600	1	33	– Челябинск	9600	1
9	– Новосибирск	9600	4	34	– Уфа	9600	1
10	– Иркутск	19200+9600	2	35	Ростов – Ташкент	9600	1
11	– Донецк	9600	1	36	– Новосибирск	9600	1
12	– Красноярск	9600	3	37	– Красноярск	9600	1
13	– Алма-Ата	9600	1	36	– Краснодар	9600	2
14	– Хабаровск	64К+19200	2	39	– Тюмень	9600	1
15	– Краснодар	9600	1	40	– Волгоград	9600	1
16	– Сочи	9600	1	41	– Якутск	9600	1
17	– Тюмень	9600	4	42	– Мин. Воды	9600	1
18	– Волгоград	9600	1	43	Иркутск – Братск	9600	1
19	– Казань	64К + 19200	2	44	Алма-Ата – Бишкек	9600	1
20	– Братск	9600	1	45	Хабаровск – Владивосток	9600	1
21	– Якутск	19200 + 9600	2	46	– Петропавловск-К.	9600	1
22	– Бишкек	9600	1	47	Волгоград – Уфа	9600	1
23	– Минеральные Воды	19200 + 9600	2	48	Владивосток – Петропавловск-К.	9600	1
24	– Уфа	9600	1				
25	– Петропавловск-Камчатский	9600	1				

В результате экспериментов были построены функциональные зависимости средней задержки сообщений по сети от величины нагрузки. Графики данных зависимостей приведены на рис. 4.2. Результаты экспериментов показывают, что:

1) средняя задержка сообщений уменьшилась по сравнению с вариантом 1989 г. (3.83 сек – в 1989 г.; 1.97 сек – в 2001 г.);

2) перегрузка сети наступает при двухкратном увеличении нагрузки на сеть для обоих вариантов.

Значительное снижение средней задержки объясняется увеличением пропускной способности каналов связи: в 1989 г. использовались, в основном, каналы с пропускной способностью 1200 – 2400 бит/сек, тогда, как в настоящее время используются каналы с пропускной способностью 9.6 Кбит/сек и выше. В то же время, использование новых технологий продажи и бронирования билетов в системах “Сирена-2000”, “Сирена-2.3” и “Сирена-3” привело к значительному увеличению потоков в линиях связи, чем объясняется относительно невысокие возможности существующей СПД “Сирена” при повышении нагрузки на сеть. Таким образом, при увеличении объема продаж авиабилетов и внедрении новых технологических функций, которые приведут к росту информационных потоков, необходимо

дальнейшая модернизация сети передачи данных “Сирена” с целью повышения пропускных способностей каналов связи.

СПИСОК ЛИТЕРАТУРЫ

1. Клейнрок Л. *Коммуникационные сети*. М.: Наука, 1975.
2. Клейнрок Л. *Вычислительные системы с очередями*. М.: Мир, 1979.
3. Шварц М. *Сети ЭВМ. Анализ и проектирование*. М.: Радио и связь, 1981.
4. Бертсекас Д., Галлагер Р. *Сети передачи данных*. М.: Мир, 1989.
5. Дэвис Д., Барбер Д., Прайс У., Соломонидес С. *Вычислительные сети и сетевые протоколы*. М.: Мир, 1981.
6. Мизин И.А., Богатырев В.А., Кулешов А.П. *Сети коммутации пакетов*. М.: Радио и связь, 1986.
7. Jackson J.R. Networks of Waiting Lines. *Operations Research*, 1957, no. 5, pp. 518–521.
8. Гнеденко Б.В., Коваленко И.Н. *Введение в теорию массового обслуживания*. М.: Наука, 1987.
9. Fratta L., Gerla M., Kleinrock L. The Flow Deviation Method: An Approach to Store-and-Forward Communication Network Design. *Networks*, 1973, vol. 3, no. 2, pp. 97–133.
10. Cantor D.G, Gerla M. Optimal Routing in a Packet-Switched Computer Network. *IEEE Trans. Computers*, 1974, vol. C-23, no. 10, pp. 1062–1068.
11. Schwartz M., Cheung C.K. The Gradient Projection Algorithm for Multiple Routing in Message-Switched Networks. *IEEE Trans. Commun.*, 1976, vol. COM-24, no. 4, pp. 449–456.
12. Gerla M., Kleinrock L. On the Topological Design of Distributed Computer Networks. *IEEE Trans. Commun.*, 1977, vol. COM-25, no. 1, pp. 48–53.
13. Frank M., Wolfe P. An Algorithm for Quadratic Programming. *Naval Research Logistic Quarterly*, 1956, no. 3, pp. 95–110.
14. Floyd R.W. Algorithm 97: Shortest Path. *Comm. ACM*, 1962, no. 3, p. 345.
15. Федотов Е.В. Определение оптимальных маршрутов в сети пакетной коммутации. В сборнике: *Сетевая обработка информации*. М.: МДНТП, 1990, стр. 95–98.
16. Gavish B., Hantler S.L. An Algorithm for Optimal Route Selection in SNA Networks. *IEEE Trans. Commun.*, 1983, vol. COM-31, no. 10, pp. 1154–1161.
17. Courtois P.J., Semal P. An Algorithm for the Optimization of Nonbifurcated Flows in Computer Communication Networks. *Performance Evaluation*, 1981, vol. 1, pp. 139–152.
18. Вишневский В.М., Федотов Е.В. Анализ методов маршрутизации при проектировании сетей пакетной коммутации. *3rd I.S. “Teletraffic Theory and Computing Modeling”*, София, 1990.
19. Мину М. *Математическое программирование. Теория и алгоритмы*. М.: Наука, 1990.
20. Поляк Б.Т. *Введение в оптимизацию*. М.: Наука, 1983.
21. Held M., Wolfe P., Growder H.P. Validation of Subgradient Optimization. *Mathematical programming*, 1974, no. 6, pp. 62–88.
22. Dijkstra E.W. A Note on Two Problems in Connection with Graphs. *Numer. Math.*, 1959, no. 1, pp. 269–271.
23. Зайченко Ю.П. Задачи проектирования структуры распределенных вычислительных сетей. *Автоматика*, 1981, № 3, стр. 35–44.
24. Жожикашвили В.А., Вишневский В.М. *Сети массового обслуживания. Теория и применение к сетям ЭВМ*. М.: Радио и связь, 1988.

Статью представил к публикации член редколлегии В.И.Венец