

Б. С. Гольдштейн

МЕТОД ОПТИМИЗАЦИИ ТЕЛЕФОННОЙ ОПЕРАЦИОННОЙ СИСТЕМЫ ЭЛЕКТРОННОГО КОММУТАЦИОННОГО УЗЛА

Операционная система (ОС) электронной управляющей машины (ЭУМ) телефонного узла коммутации – телефонная операционная система (ТОС) – представляет собой упорядоченную совокупность программных средств, организующих процесс обслуживания вызовов в реальном масштабе времени, т. е. в естественном темпе функционирования узла коммутации (УК), и обеспечивающих управление программными и аппаратными ресурсами ЭУМ. Наиболее дефицитным аппаратным ресурсом ЭУМ является время работы его центрального процессора (ЦП). Оптимальное распределение этого времени между отдельными функциональными программами – один из главных вопросов, возникающих при проектировании ТОС.

В статье предлагается метод оптимизации функционирования ТОС путем распределения приоритетов функциональных программ и оценивается эффективность работы такой ТОС. Архитектура телефонной операционной системы электронного узла коммутации рассмотрена в [1].

Работа программного обеспечения ЭУМ, взаимодействие программных блоков ТОС осуществляются по системе приоритетного обслуживания [1], которая позволяет определять, в какой последовательности, когда и какая из телефонных программ должна выполняться в конкретный момент времени. Исследования показали, что для ЭУМ коммутационных узлов и станций эффективными являются дисциплины обслуживания со смешанными абсолютно-относительными приоритетами [1].

Для описания этой дисциплины представим приоритеты в виде пары чисел (k, m) подобно тому, как это было сделано в [2]. Такое представление приоритетов позволяет описать любую стратегию обслуживания вызовов данной ЭУМ УК. Например, имеется N потоков запросов на включение N телефонных программ, которым поставлены в соответствие N приоритетов. Пусть эти N приоритетов распределены каким-то образом по K уровням абсолютного приоритета. Обслуживание запроса любого приоритета уровня $k \in \{2, F\}$ прерывается при появлении другого запроса, приоритет которого соответствует меньшему значению $k = \{1, K - 1\}$. На каждом уровне k располагается M_k приоритетов, запросы которых не прерывают друг друга и взаимодействуют по принципу относительных приоритетов. Запросы, имеющие одинаковый приоритет, обслуживаются по правилу «первым пришел – первым обслуживается».

Тогда приоритеты можно описывать в виде пары чисел (k, m) . На уровне k таких пар будет $1 \dots M_k$. Очевидно, $\sum_{k=1}^K M_k = N$. Приоритетная стратегия ТОС однозначно определяется разбиением вида $\mu = \{M_1, M_2, \dots, M_k\}$.

На рис. 1 показана организация приоритетного обслуживания запросов в телефонной операционной системе. Здесь a – потоки запросов, прерывающих функциональную программу, которая обслуживает запрос с приоритетом (k, m) ; b – потоки запросов, не прерывающих функциональную программу, которая обслуживает запрос с приоритетом

(k, m) ; ν – потоки запросов, обслуживание которых не прерывается при появлении запроса с приоритетом (k, m) ; ε – потоки запросов, обслуживание которых прерывается при появлении запроса с приоритетом (k, m) .

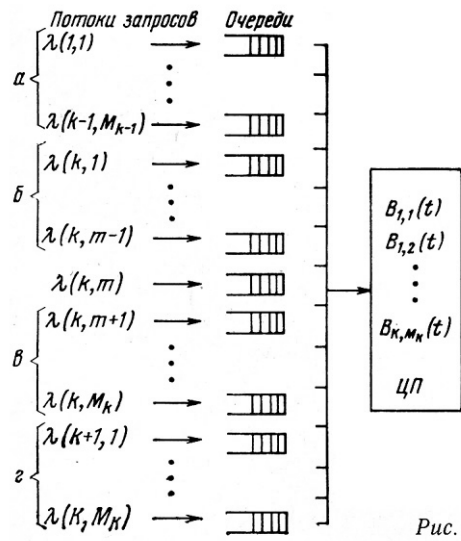


Рис. 1

Аналитическое исследование рассмотренной математической модели приоритетной стратегии ТОС становится возможным при введении следующих допущений. Потоки запросов на включение программ являются независимыми пуассоновскими с параметрами $\lambda(k, m)$. Время выполнения программы, обслуживающей потоки запросов различных приоритетов, – независимая случайная величина с функцией распределения $B_{k,m}(t)$, математическим ожиданием $b(k, m)$ и конечным вторым начальным моментом $b^2(k, m)$. Размеры очередей запросов выбираются таким образом, чтобы исключить возможность потери запроса из-за отсутствия мест для ожидания; прерывание обслуживающей программы не приводит к потере уже затраченного на ее обслуживание времени.

Рассматривается стационарный режим работы системы, т.е. суммарная загрузка ТОС

$$R = \sum_{i=1}^K \sum_{j=1}^{M_i} \lambda(i, j) b(i, j) = \sum_{i=1}^K \sum_{j=1}^{M_i} \rho(i, j) < 1. \tag{1}$$

Аналитические выражения для расчета математического ожидания времени пребывания запросов в ЭУМ, полученные в статье, несколько отличаются от приведенных в [3, с. 120–122]. Очевидно, что математическое ожидание времени $T(k, m)$ пребывания запросов с приоритетом (k, m) в ЭУМ складывается из среднего полного времени обслуживания запроса с приоритетом (k, m) – $V(k, m)$ и среднего времени ожидания начала работы программы, обслуживающей запрос с приоритетом (k, m) , – $W(k, m)$.

Среднее полное время обслуживания запроса с приоритетом (k, m) , т. е. время от начала выполнения программы до полного завершения обслуживания запроса,

$$V(k, m) = \frac{b(k, m) - \theta(k, m) \sum_{j=1}^{m-1} \rho(k, j)}{1 - \sum_{i=1}^k \sum_{j=1}^{\varphi(k, m-1)} \rho(i, j)}, \tag{2}$$

где $\theta(k, m)$ – математическое ожидание времени от последнего прерывания в работе обслуживающей программы до завершения обслуживания запроса с приоритетом (k, m) ,

$$\theta(k, m) = \begin{cases} b(k, m) & \text{ïðå } k = 1; \\ \frac{1}{\sum_{i=1}^{k-1} \sum_{j=1}^{M_i} \lambda(i, j)} \int_0^{\infty} \left[1 - \exp \left\{ -t \sum_{i=1}^{k-1} \sum_{j=1}^{M_i} \lambda(i, j) \right\} \right] d B_{k, m}(t) & \text{ïðå } k \neq 1. \end{cases} \quad (3)$$

$$W(k, m) = \frac{\frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{\varphi(k, m)} \lambda(i, j) b^{(2)}(i, j) + \left[1 - \sum_{i=1}^{k-1} \sum_{j=1}^{M_i} \rho(i, j) \right] \sum_{j=m+1}^{M_k} \rho(k, j) v(k, j)}{\left[1 - \sum_{i=1}^k \sum_{j=1}^{\varphi(k, m-1)} \rho(i, j) \right] \times \left[1 - \sum_{i=1}^k \sum_{j=1}^{\varphi(k, m)} \rho(i, j) \right]}, \quad (4)$$

где $v(k, j)$ – математическое ожидание времени работы программы, обслуживающей запрос с приоритетом (k, j) при наличии в ЭУМ запроса с приоритетом (k, m) при $m < j$;

$$v(k, j) = \begin{cases} \frac{b^{(2)}(k, j)}{2b(k, j)} & \text{ïðå } k = 1; \\ \frac{b(k, j) - \theta(k, j)}{b(k, j) \sum_{i=1}^{k-1} \sum_{j=1}^{M_i} \lambda(i, j)} & \text{ïðå } k \neq 1. \end{cases} \quad (5)$$

Очевидно, что математическое ожидание времени пребывания в ЭУМ запросов с приоритетом (k, m)

$$T(k, m) = W(k, m) + V(k, m). \quad (6)$$

Выражения (2)...(6) позволяют производить априорную оценку временных характеристик функционирования ТОС. Следует заметить, что полученные формулы справедливы и для систем с «чистыми» абсолютными или относительными приоритетами. Действительно, при $K=N$ и $M_i=1, \forall i$, имеет место обслуживание с абсолютными приоритетами, а при $K=1$ и $M_i=N$ – с относительными приоритетами. Инженерные расчеты $T(k, m)$ по предложенной методике выполняются с помощью пакета прикладных программ расчета параметров ТОС . PATOS [4]. Именно так были построены графики зависимости математического ожидания времени пребывания запросов в ЭУМ T от номера приоритета l для приоритетной стратегии $\mu = \{3, 2, 5, 6, 4\}$ при различных значениях суммарной загрузки ЭУМ R , приведенные на рис. 2 при $N = 20$. Полагаем, что $\lambda(i, j) = \lambda b(i, j) = b$ для $\forall i, j$. Основным преимуществом организации ТОС с приоритетами является сохранение почти неизменным среднего времени пребывания в ЭУМ запросов с высшими приоритетами при увеличении суммарной загрузки ЭУМ.

Организация телефонной операционной системы с приоритетами позволяет адаптировать процесс управления УК к изменениям интенсивности операционного трафика, например при резком возрастании суммарной загрузки ЭУМ существенно увеличивается пропускная способность ЭУМ за счет некоторого кратковременного снижения надежности управления и замедления взаимодействия ЭУМ с эксплуатационным персоналом.

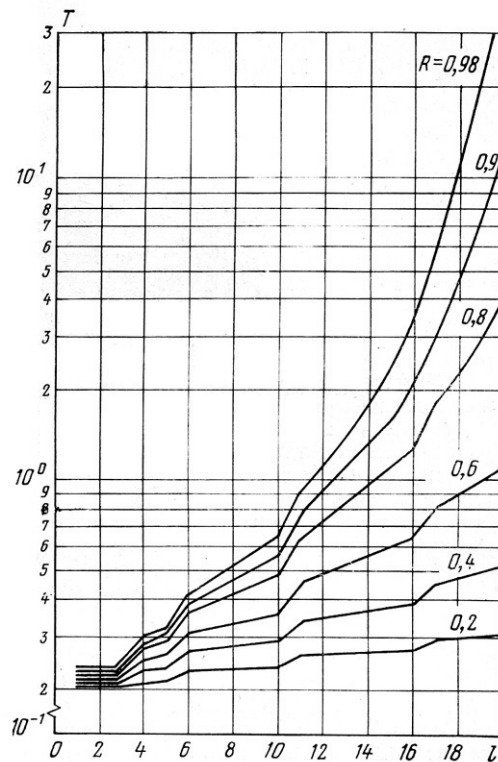


Рис. 2

Имеются некоторые непроизводительные затраты времени процессора и оперативной памяти ЭУМ, которые связаны с сохранением телефонной программы при ее прерывании и с запоминанием содержимого регистра адреса текущей команды, соответствующих признаков, содержимого базовых и операционных регистров и т. д. для последующего дообслуживания прерванной программы. Эти затраты ресурсов ЭУМ, а также объем аппаратных средств на реализацию системы прерываний пропорциональны числу уровней абсолютного приоритета K .

Тогда критерий оптимальности приоритетной организации ТОС можно определить следующим образом: при заданном наборе функциональных программ узла коммутации и характеристик ЭУМ приоритетная стратегия, определяемая разбиением $\mu = \{M_1, M_2, \dots, M_K\}$, является оптимальной, если обеспечивается выполнение всех временных ограничений вида

$$\{T(i, j) \leq \tau(i, j)\} \quad i = \overline{1, K}, j = \overline{1, M_K}, \quad (7)$$

при минимальном значении числа уровней абсолютного приоритета K .

Рассмотрим порядок определения допустимого времени пребывания запросов в ЭУМ τ . Величина τ зависит от заданного качества функционирования коммутационного узла и структуры его программного обеспечения (порядка разбиения телефонных алгоритмов на программные модули и количества последних).

Обозначим имеющийся набор программных модулей через $P = \{p_1, p_2, \dots, p_N\}$. Требуется определить некоторую перестановку этого набора $P_I = \{p_{i1}, p_{i2}, \dots, p_{iN}\}$, задающую порядок возрастания значений допустимого времени обслуживания запросов на включение этих программ, и сами эти значения.

На множестве программ P можно ввести отношение предшествования \succsim . Запись $p_i \succsim p_j$ равносильна утверждению, что программа p_j получает управление или информацию от

программы p_i . Отношение транзитивно, т. е. $\forall p_i, p_j, p_k, p_i \sim p_j, p_j \sim p_k \Rightarrow p_i \sim p_k$. Отношение \sim не является симметричным и рефлексивным. Две программы независимы, если p_i не $\sim p_j$ и p_j не $\sim p_i$.

Для множества программ узла коммутации имеется целый ряд временных ограничений – допустимых интервалов $t_{\text{доп}}$ – на выполнение определенных последовательностей программ. Эти ограничения связаны с изменениями состояний оборудования узла, с принятой системой обмена сигналами управлений и взаимодействия с соседними узлами и станциями, с порядком взаимодействия с абонентами данного узла. К числу таких допустимых интервалов относятся время от момента снятия абонентом телефонной трубки до отправки сигнала «Готов к приему номера», время между отсылками отдельных цифр номера (например, 500 мс для связи с АТСДШ), время между отсылкой отдельных сигналов управления (например, 45 ± 5 мс при работе с АТСК), интервалы отсчета тарификационных импульсов, время между импульсами набора номера (10 мс при передаче цифр номера к АТСДШ) и др.

Некоторые допустимые временные интервалы $t_{\text{доп}}$ определяются эмпирически. Например, в ЭУМ предусмотрена возможность получения ответов на запросы эксплуатационного персонала узла коммутации. Для этого в состав оборудования узла обычно включается пульт оператора, для организации работы которого разрабатываются специальные программы. Работа этих программ считается удовлетворительной, если время ответа ЭУМ на запрос эксплуатационного персонала с пульта оператора не слишком велико по сравнению с монопольным временем ответа ЭУМ, т. е. временем обслуживания только данного запроса оператора. Количественно этот критерий может быть сформулирован следующим образом: время ответа ЭУМ на запрос оператора в ЧНН не должно превышать двадцатикратное монопольное время ответа ЭУМ при длительном диалоге оператора с ЭУМ или 5с при кратком диалоге.

Очевидно, что задаваемые допустимые временные интервалы $t_{\text{доп}}$ относятся к программам, связанным отношением предшествования $p_i \sim p_j$. При этом, как правило, эти программы не связаны отношением непосредственного предшествования, т. е. $\exists p_s, p_i \sim p_s \sim p_j$. Тогда задача определения допустимых времен пребывания запросов в ЭУМ сводится к отысканию максимальных значений τ_i , удовлетворяющих всем неравенствам вида

$$\sum_{j=1}^N c_{k,l}^i \tau_j \leq t_{\text{доп}}^i(k,l). \quad (8)$$

где

$$c_{k,l}^i = \begin{cases} 1, & \text{если выполняются отношения } p_k \sim p_i \text{ и } p_i \sim p_l; \\ 0, & \text{если хотя бы одного такого отношения не существует.} \end{cases}$$

Так как число потоков запросов N на включение функциональных программ конечно, задача оптимизации приоритетной ТОС может быть решена простым перебором и выбором лучшего из всех возможных вариантов разбиения $\mu = \{M_1, M_2, \dots, M_k\}$ при условии выполнения ограничений (7) хотя бы для одного такого варианта. Однако нетрудно показать, что общее число возможных комбинаций равно 2^{N-1} и при больших N метод перебора нереален.

Алгоритм направленной оптимизации приоритетной стратегии ТОС, позволяющий получить решение при рассмотрении малого числа возможных вариантов разбиений, приведен на рис. 3. Поскольку в алгоритме осуществляется направленный перебор конечного числа вариантов разбиений и на каждом шаге отбрасываются недопустимые

или неперспективные (в смысле минимизации K) варианты, сходимость алгоритма очевидна.

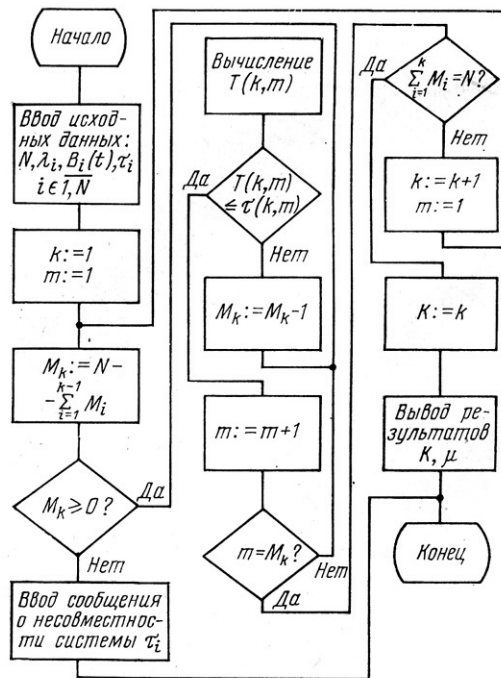


Рис. 3

Благодаря оптимизации ТОС по такому алгоритму процессор обслуживает запросы на включение телефонных программ, соблюдая все временные соотношения с минимальным запасом, стараясь при этом занять наименьший объем оперативной памяти на запоминание прерванных задач и ограничить время процессора на обработку сигналов прерывания.

Рассмотрим работу алгоритма оптимизации ТОС на конкретном примере. Исходные данные следующие: число потоков запросов $N=9$; допустимое время пребывания запросов в ЭУМ $\tau_1=\tau_2=0,3$; $\tau_3=0,4$; $\tau_4=0,6$; $\tau_5=0,65$; $\tau_6=\tau_7=2,5$; $\tau_8=\tau_9=20,0$. Суммарная нагрузка ЭУМ $R = 0,9$. Полагаем, что интенсивности всех потоков запросов одинаковы и равны $\lambda = 0,1$, а математические ожидания времени работы функциональных программ, обслуживающих эти запросы, также одинаковы и равны $b = 0,2$.

Результаты вычисления $T(k, m)$ по данному алгоритму, сведем в таблицы. Здесь значком * отмечены значения $T(k, m)$, для которых проверяется их соответствие заданным значениям $\tau_1 \dots \tau_9$; при их превышении рассчитывается $T(k, m)$ для следующего значения $M_k = M_k - 1$; № – порядковый номер потока запросов; k, m – присваиваемый приоритет. При $k=1$

№	1	2	3	4	5	6	7	8	9
k, m	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9
$T(k, m)$	0,4* 0,378* 0,356* 0,333* 0,311* 0,289* 0,267	0,45 0,422 0,394 0,367 0,339 0,311* 0,283*	0,521 0,486 0,45 0,414 0,379 0,343 0,307*	0,629 0,581 0,533 0,486 0,438 0,391	0,8 0,733 0,667 0,6 0,533	1,1 1,0 0,9 0,8	1,7 1,533 1,367	3,2 2,867	9,2

Тогда $M_1=3$. При $k = 2$:

№	4	5	6	7	8	9
k, m	2,1	2,2	2,3	2,4	2,5	2,6
$T(k, m)$	0,604* 0,579* 0,553 0,528 0,502	0,785 0,749* 0,713* 0,677* 0,641*	1,1 1,046 0,992 0,939	1,731 1,641 1,551	3,308 3,128	9,615

Тогда $M=2$. При $k = 3$:

№	6	7	8	9
k, m	3,1	3,2	3,3	3,4
$T(k, m)$	1,1*	1,744*	3,356*	9,8*

Тогда $M_3=4$, $K=3$ и работа алгоритма заканчивается.

Таким образом, для девяти потоков запросов ($N=9$) оптимальная приоритетная стратегия ТОС, определяемая разбиением $\mu=\{3, 2, 4\}$, была найдена через 13 циклов алгоритма просчетом 13 вариантов. Методом полного перебора потребовалось бы просчитать 256 вариантов. Это подтверждает высокую эффективность предлагаемого алгоритма, которая возрастает с увеличением N .

С помощью данного алгоритма разработаны три версии телефонной операционной системы для коммутационных узлов различного типа, управляемых ЭУМ семейства «Нева» [6]. Опытная эксплуатация этих ТОС показала, что процесс обслуживания вызовов в них организован наиболее эффективно с точки зрения как максимального использования ресурсов ЭУМ, так и качества функционирования коммутационного узла.

ЛИТЕРАТУРА

1. Гольдштейн Б.С. Телефонная операционная система электронного узла коммутации. – Электросвязь, 1980, № 8.
2. Herzog U. Optimal scheduling strategies for real time computers. – IBM Journal of Research and Development, 1975, September.
3. Балыбердин В.А. Методы анализа мультипрограммных систем / Под ред. С.Д. Пашкеева. – М.: Радио и связь 1982.
4. Бойцова Л.В., Гольдштейн Б.С. Алгоритм расчета основных параметров комбинированной системы приоритетного обслуживания. – Информационный бюллетень Госфонда СССР. Алгоритмы и программы, 1979, № 5 (31).
5. Гольдштейн Б.С. Тенденции развития операционных систем коммутационных узлов и станций с программным управлением. – Экспресс-информация. Зарубежная техника связи. Серия: Телефония. Телеграфия. Передача данных. – М.: ЦНТИ Информсвязь; 1980, вып. 11.
6. Бескинд А.А., Шляпоберский В.И., Штагер В.В. Построение и использование управляющих комплексов семейства «Нева». – Электросвязь, 1979, № 7.

Статья поступила 27 декабря 1983 г.