МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра вычислительной техники

УТВЕРЖДАЮ
Проректор по учебной работе
О.Г. Локтионова
«30» 09 2022 г.

Математические методы построения инфокоммуникационных сетей и систем

Методические указания к практическим работам
для студентов направления подготовки
09.04.01 очной формы обучения

Курск 2022

УДК 001.89

Составители: Д.В. Быков, А.В. Киселев, Е.А. Кулешова

Рецензент
Кандидат технических наук, доцент *Т.Н. Конаныхина*

Математические методы построения инфокоммуникационных сетей и систем: методические указания к практическим работам для студентов направления подготовки 09.04.01 очной формы обучения / Юго-Зап. гос. ун-т; сост.; Д.В. Быков, А.В. Киселев, Е.А. Кулешова. – Курск, 2022. - 47 с.: - ил. 13, табл. 3.– Библиогр.: с. 47.

В методических указаниях рассмотрены основные положения теории сетей массового обслуживания. Представлен расчет замкнутой и открытой сети массового обслуживания.

Предназначены для студентов направления подготовки 09.04.01 очной формы обучения.

Методические указания соответствуют рабочей программе дисциплины «Математические методы построения инфокоммуникационных сетей и систем».

Текст печатается в авторской редакции

Подписано в печать . Формат 60∗84 1/16.
Усл. печ. л. 2,85. Уч.-изд. л. 2,58. Тираж 50 экз. Заказ . Бесплатно.
Юго-Западный государственный университет.
305040 Курск, ул. 50 лет Октября, 94.

1. РАСЧЕТ СЕТЕЙ МАССОВОГО ОБСЛУЖИВАНИЯ

1.1. Основные положения теории сетей массового обслуживания

Сеть массового обслуживания (СеМО) представляет собой совокупность конечного числа центров обслуживания, в которой циркулируют заявки, переходящие в соответствии с маршрутной матрицей из одного центра в другой. Центры сети являются системами массового обслуживания и отображают функционально самостоятельные части моделируемой системы, а связи между ними − структуру системы. В общем случае, центр состоит из A одинаковых обслуживающих приборов ($0 \le A \le \infty$) и буфера (очереди ожидания) объемом C ($0 \le C \le \infty$).

Переход заявки из одного центра после окончания обслуживания в нем в другой осуществляется в соответствии с заданным маршрутом, под которым понимается последовательность посещаемых заявкой центров сети. Маршрут заявки по СеМО задается матрицей переходных вероятностей P, вид которой зависит от того, является ли сеть МО открытой или замкнутой. В открытую сеть сообщения поступают из внешнего источника и могут покинуть сеть после окончания обслуживания. В замкнутой сети число сообщений не меняется, они не покидают сеть. Если принять внешний источник за новый центр сети и обозначить индексом 0, то маршрут в открытой сети задается стохастической неразложимой матрицей $P=\|P_{ij}\|$, где P_{0j} и P_{j0} – соответственно вероятность поступления в j–й центр заявки из источника и вероятность покидания заявкой сети после окончания обслуживания в j–ом центре; P_{ij} – вероятность того, что заявка, уходящая из i-го центра, перейдет в j-й центр ($i, j = \overline{1,R}$), R – количество центров сети. Очевидно, что выполняется равенство:

$$\sum_{j=0}^{R} P_{ij} = 1,\ i = \overline{0,R},\ P_{00} = 0 \qquad (1.1)$$

Описание первых подходов к анализу сетей массового обслуживания было дано в работах Джексона, где были рассмотрены однородные разомкнутые СеМО с пуассоновскими входящими потоками, экспоненциальными распределениями

длительности обслуживания в центрах и дисциплинами обслуживания FCFS (в порядке поступления) на узлах типа M/M/m.

Функционирование такой сети описывается процессом гибели и размножения. В результате решения системы линейных уравнений, называемых уравнениями баланса, получено равновесное совместное распределение количества заявок в центрах $P(n_1, n_2, ..., n_R)$ (т. е. вероятность того, что в первом центре находится n_1 заявок, на втором — n_2 и т. д., всего R узлов) в виде произведения маргинальных распределений, известное под названием теоремы разложения Джексона:

$$P(n_1, n_2, ..., n_R) = \prod_{i=1}^{R} P_i(n_i), \qquad (1.2)$$

где $P_i(n_i)$ – стационарная вероятность того, что в i-м центре, рассматриваемом изолированно, находится n_i сообщений.

Хорошо известно обобщение Гордона и Ньюэла результата (1.2) на случай **однородных замкнутых** сетей, которые отличаются от разомкнутых тем, что в них отсутствуют внешние поступления заявок и уходы их из сети.

В этом случае вид формулы (1.2) остается прежним, но в нее должна входить нормировочная константа G, которая должна обеспечивать выполнение равенства единице суммы вероятностей состояний сети при условии, что сеть содержит ровно N заявок:

$$P(n_1, n_2, ..., n_R) = \frac{1}{G} \prod_{i=1}^{R} Z_i(n_i), \qquad (1.3)$$

где введено обозначение:

$$Z_i(n_i) = \frac{e_i^{n_i}}{\prod_{j=1}^{n_i} \mu_i(j)}, \qquad (1.4)$$

Выражения (1.2), (1.3) позволяют получить для открытых и замкнутых экспоненциальных сетей решение в мультипликативной форме, допускающей декомпозицию сети на изолированные центры.

Для определения потоков, циркулирующих в стационарном режиме в

открытой сети МО, введем коэффициенты передачи e_i, такие, что $\lambda(N)e_i$ представляет собой общую интенсивность потока сообщений в i-й центр сети $(i=\overline{1,R})$:

$$\lambda_i(N) = e_i\lambda(N),\ i=\overline{1,R} \qquad (1.5)$$

В **открытых** СеМО интенсивность λ_i складывается из интенсивности поступления сообщений в i-й центр из источника и интенсивности поступления из других центров:

$$e_i = P_{oi} + \sum_{j=1}^{R} P_{ji}e_j\ ,\ i=\overline{1,R} \qquad (1.6)$$

Для **замкнутых** сетей исключается поток от внешнего источника и система уравнений (1.6) преобразуется к виду

$$e_i = \sum_{j=1}^{R} P_{ji}e_j\ ,\ i=\overline{1,R} \qquad (1.7)$$

Для отыскания однозначного решения системы уравнений (1.7) достаточно произвольно задать значение e_i, например положить $e_1=1$. В этом случае величину e_i можно интерпретировать как среднее число посещений центра i между двумя последовательными посещениями им первого центра.

1.2 Расчет замкнутой сети массового обслуживания

Предположения об экспоненциальном распределении времени обслуживания в центрах и о дисциплине FCFS в реальных системах выполняются далеко не всегда. Однако, в ряде случаев, когда порядок обслуживания отличен от FCFS, а распределение длительности обслуживания не является экспоненциальным, основной результат Джексона (1.2) остается справедливым.

В соответствии с результатом, известным как теорема BCMP, мультипликативное свойство решений (1.2) и (1.3) для $P(n_1,n_2,...,n_R)$ сохраняется для СеМО, содержащих следующие узлы:

а) M/M/m с дисциплиной FCFS;

б) M/G/1 с дисциплиной PS (разделение процессора);

в) M/G/∞ с обслуживанием без ожидания (IS узлы);

г) M/G/l с дисциплиной LCFS с прерываниями.

Указанные системы обслуживания обладают тем свойством, что выходящий поток обслуженных требований в стационарном режиме является пуассоновским. Для систем а) и б) этот факт установлен Бэрком, Рейчем и Дубом.

При этом множители $Z_i(n_i)$, входящие в (1.3), имеют вид:

$$Z_i(n_i) = \frac{n_i!}{\mu_i^{n_i}} \prod_{h=1}^{H} \frac{e_{ih}^{n_{ih}}}{n_{ih}} \text{ для центров с дисциплиной FCFS,} \quad (1.8)$$

$$Z_i(n_i) = \prod_{h=1}^{H} \frac{1}{n_{ih}!} \left(\frac{e_{ih}}{\mu_{ih}} \right)^{n_{ih}} \text{ для центров с дисциплиной IS,} \quad (1.9)$$

$$Z_i(n_i) = n_i! \prod_{h=1}^{H} \frac{1}{n_{ih}!} \left(\frac{e_{ih}}{\mu_{ih}} \right)^{n_{ih}} \text{ для центров с дисциплинами PS и LCFC,} \quad (1.10)$$

где $n_i = \sum_{h=1}^{H} n_{ih}$, $i = \overline{1, R}$, H – количество классов сообщений, R - количество центров сети.

Если нормировочная константа G определена, то не составляет трудности получение основных характеристик СеМО – маргинальных распределений числа заявок в центрах, средних длин очередей и времен ожиданий, пропускной способности сети и т.д.

1.1.1 Метод Бузена

Трудоемкость вычисления нормировочной константы пропорциональна числу узлов R и числу требований N и в настоящее время разработан ряд алгоритмов ее расчета, сокращающих число вычислительных операций.

В основе большинства из них лежит рекуррентный метод Бузена. В

соответствии с этим алгоритмом вычисление нормировочной константы для **замкнутой** СеМО, с единственным классом сообщений, содержащей N заявок и R узлов, осуществляется с помощью рекуррентного выражения ($G_R(N) = g(N,R)$):

$$g(n,r) = \sum_{k=0}^{n} Z_r(k) g(n-k, r-1), \qquad (1.11)$$

где

$$Z_i(k) = \begin{cases} \dfrac{e_i}{\mu_i(k)} Z_i(k-1), & \text{если центр } i \text{ зависит от нагрузки,} \\ \dfrac{e_i}{\mu_i} Z_i(k-1), & \text{если центр } i \text{ не зависит от нагрузки,} \end{cases}$$

и начальными условиями $Z_i(0) = 1$, $g(n, 1) = Z_1(n)$ $g(0,r) = 1$ для $k, n = \overline{1, N}$ и $i, r = \overline{1,R}$.

Интенсивность обслуживания центра i, **зависящего от нагрузки**, определяется количеством занятых обслуживающих устройств:

$$\begin{gathered} \mu_i(n) = n\mu_i, \text{ если } 0 \le n \le A_i,\ n = \overline{1, N},\ i = \overline{1, R}, \\ \mu_i(n) = A_i\mu_i, \text{ если } n > A_i,\ n = \overline{1, N},\ i = \overline{1, R}, \end{gathered} \qquad (1.12)$$

где A_i – общее количество обслуживающих устройств i-го центра.

Для центра, **не зависящего от нагрузки**, выражение (1.11) упрощается:

$$g(n,r) = g(n,r-1) + x_r g(n-1,r), \qquad (1.13)$$

где $x_r = \dfrac{e_r}{\mu_r}$, начальные условия $g(0, r)=1$, $g(n, 0)=0$ и $n = \overline{1, N}$, $r = \overline{1, R}$.

В результате рекуррентных вычислений (1.11), (1.13) получается таблица (столбцы – центры обслуживания, строки – количество сообщений в сети), последний элемент последнего столбца которой представляет собой искомую нормировочную константу $G_R(N)$.

Таблица 1.1 Таблица расчета нормировочной константы по методу Бузена

	1	2	…	r-1	r	…	R
0	1	1	…	1	1	…	1
1	x_1						$G_R(1)$
…	…		…				…
n-1	x_1^{n-1}				$g(n-1, r)$		$G_R(n-1)$
n	x_1^{n}			$g(n,r-1)$	$g(n,r)$		$G_R(n)$
…							…
N	x_1^{N}						$G_R(N)$

Для сети, **центры которой зависят от нагрузки**, маргинальное распределение числа заявок в R-м (граничном центре) определяется в виде

$$P_R(n,N)=\frac{Z_R(n)G_{R-1}(N-n)}{G_R(N)},\ n=\overline{1,N},\ i=\overline{1,R}. \qquad (1.14)$$

Маргинальное распределение в любом центре обслуживания $i \neq R$ может быть получено путем перенумерации центров так, чтобы центр, представляющий интерес, стал граничным.

Интенсивность выходящего потока заявок (пропускная способность) из i-го центра $\lambda_i(N)$ для $i=\overline{1,\ R}$

$$\lambda_i=\frac{e_i G_i(N-1)}{G_i(N)}. \qquad (1.15)$$

Математическое ожидание числа сообщений в R-м центре имеет вид

$$S_R(N)=\frac{1}{G_R(N)}\sum_{n=1}^{N} nZ_R(n)G_{R-1}(N-n). \qquad (1.16)$$

В соответствии с формулой Литтла среднее время пребывания заявки в i–м центре обслуживания $T_i(N)$ равно отношению математического ожидания числа сообщений в центре к средней интенсивности входящего потока. В стационарном режиме интенсивность выходящего потока равна интенсивности входящего, поэтому

$$T_i(N) = S_i(N)/\lambda_i(N),\ i = \overline{1,\ R}. \qquad (1.17)$$

Если сеть не зависит от нагрузки или зависимость интенсивности обслуживания центров в ней определяется из соотношений (1.12), то перенумерация центров не требуется, и маргинальное распределение числа заявок в любом центре определяется в виде

$$P_i(n, N) = \frac{x_i^n [G_R(N-n) - x_i G_R(N-n-1)]}{G_R(N)},\ n = \overline{1, N},\ i = \overline{1, R}. \qquad (1.18)$$

Математическое ожидание числа сообщений в *i*-м центре имеет вид

$$S_i(N) = \frac{1}{G_R(N)} \sum_{n=1}^{N} x_i^n G_M(N-n),\ i = \overline{1, R}. \qquad (1.19)$$

Вычисление нормировочной константы методом Бузена при увеличении количества заявок в СеМО резко усложняется и при этом, в случае реализации на ЭВМ, возникает проблема переполнения и обнуления результатов, кроме того, необходимо перенумеровывать центры для определения характеристик всех центров. Поэтому были предложены ряд методов, позволяющих оптимизировать процесс вычисления по методу Бузена, а также альтернативные методы расчета характеристик СеМО.

Для снижения вероятности переполнения и обнуления результатов при расчете нормировочной константы, применяется масштабирование. Выбирается масштабный коэффициент (шкалирующий множитель) C для e_i, позволяющий значительно уменьшить скорость возрастания чисел, используемых при вычислении нормировочной константы. Коэффициент C выбирается из следующих соображений: он должен минимизировать функцию $\sum_{i=1}^{R} (1 - Cx_i)^2$. Решением этой задачи является:

$$C = \frac{\sum_{i=1}^{R} x_i}{\sum_{i=1}^{R} x_i^2}, \qquad (1.20)$$

где $x_i = \frac{e_i}{\mu_i}$.

При расчете нормировочной константы и характеристик СеМО на ее основе принимается:

$$e'_i = Ce_i,\ i = \overline{1, R}. \qquad (1.21)$$

1.1.2 Метод анализа средних значений

Несмотря на оптимизацию расчетов по методу Бузена, существует ограничение на количество заявок в СеМО, при превышении которого возможны проблемы вычисления характеристик системы. Поэтому для определения таких показателей как средние длины очередей и времена пребывания в центрах, производительности и коэффициентов загрузки центров более предпочтительным является использование метода анализа средних значений (англ. – Mean Value Analysis - MVA), предложенного Рейзером и Левенбергом. В основе метода лежит закон Литтла, и он позволяет итерационно вычислять характеристики сети в зависимости от количества заявок N в СеМО. Рассмотрим расчет характеристик СеМО по алгоритму MVA.

Если обозначить однородную экспоненциальную СеМО, не зависящую от нагрузки через $D(N)$, то среднее время обработки сообщения в i-м центре с дисциплиной обслуживания FCFS складывается из средней длительности обслуживания вновь поступившего сообщения τ_i и средней длительности обслуживания всех сообщений, находящихся в i-м центре $\tau_i\ \nu_i\ (N)$:

$$T_i(N) = \tau_i\ (1 + \nu_i\ (N)),\ i = \overline{1, R}, \qquad (1.22)$$

где $\nu_i(N)$ – среднее количество сообщений в i-м центре в момент поступления

нового сообщения.

Стационарные вероятности состояний сети $D(N)$ в момент поступления сообщения в i-й центр совпадают со стационарными вероятностями состояний сети $D(N\text{-}1)$ для произвольного момента времени. Отсюда непосредственно следует, что $v_i(N) = S_i(N\text{-}1)$, где $S_i(N\text{-}1)$ – среднее число сообщений в i-м центре сети $D(N\text{-}1)$, и (1.22) приобретает вид:

$$T_i(N) = \tau_i\,(1 + S_i\,(N\text{-}1)),\ i = \overline{1, R}\,. \qquad (1.23)$$

Такая же формула может применяться для расчетов среднего время обработки сообщения в центрах с дисциплинами обслуживания LCFS и PS. Для центров с дисциплиной обслуживания IS среднее время обработки сообщения будет находиться из соотношения:

$$T_i(N) = \tau_i\,,\ i = \overline{1, R}\,. \qquad (1.24)$$

Суммарная интенсивность потока сообщений в СеМО определяется из соотношения:

$$\lambda(N) = \frac{N}{\sum_{j=1}^{R} e_j T_j(N)},\ j = \overline{1, R}\,, \qquad (1.25)$$

где e_i определяется из (1.6) или (1.7).

С учетом (1.5) интенсивность потока сообщений, входящего в i-й центр равна:

$$\lambda_i(N) = \frac{e_i N}{\sum_{j=1}^{R} e_j T_j(N)},\ i, j = \overline{1, R}\,. \qquad (1.26)$$

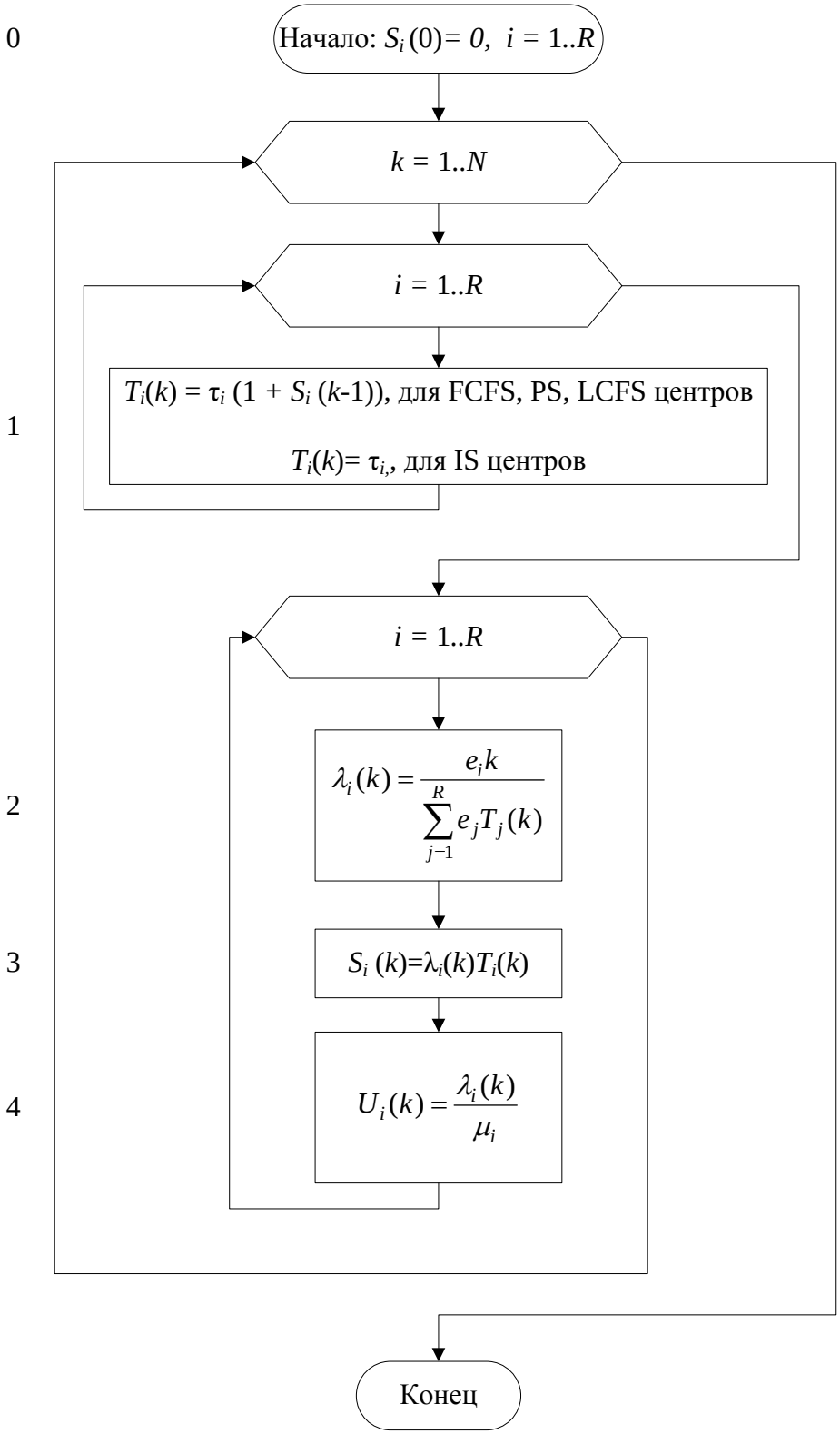

Рисунок 1.1. Реализация метода анализа средних значений для замкнутой СеМО

Используя формулу Литтла можно найти среднее количество сообщений в *i*-м центре:

$$S_i(N)=\lambda_i(N)T_i(N),\ i=\overline{1,R}, \qquad (1.27)$$

Среднее количество занятых приборов обслуживания в *i*-м центре сети равно отношению интенсивности входящего потока сообщений к интенсивности обработки сообщений (μ_i) в центре:

$$U_i(N)=\frac{\lambda_i(N)}{\mu_i},\ i=\overline{1,R}. \qquad (1.28)$$

На рисунок 1.1 приведен алгоритм MVA в общем виде.

Метод анализа средних значений позволяет также определить маргинальное распределение средней длины очереди в *i*-м центре:

$$P(n,N)=\frac{e_i\lambda(N)}{\mu_i}P_i(n-1,N-1), \qquad (1.29)$$

где $n \geq 1$, $P_i(0,0)=1$, $P_i(n,0)=0$, $i=\overline{1,R}$.

2. ПРИМЕР

2.1. Модель конфиденциального хранилища электронных документов

С учетом особенностей структуры и процессов функционирования КХЭД, его работа может быть формализована в виде разомкнутой СеМО с блокировками и отказами. Для исследования КХЭД предлагается сеть массового обслуживания, формализующая функционирование конфиденциального хранилища ЭД (рисунок 2.1).

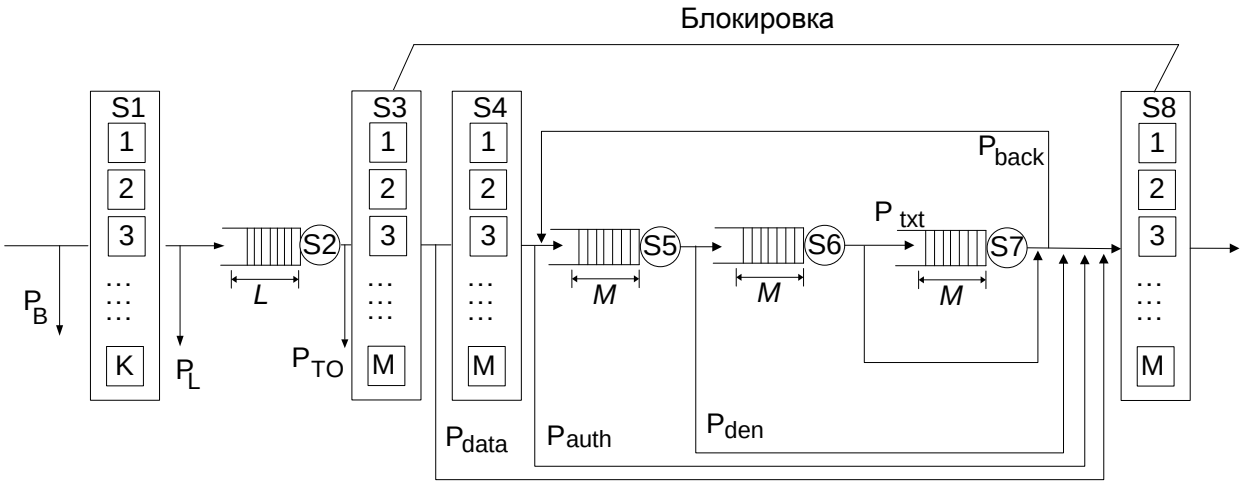

Рисунок 2.1. Сеть массового обслуживания, формализующая работу конфиденциального хранилища ЭД

S1 – центр, формализующий работу модуля TCP операционной системы сервера приложений на этапе установления соединения. K – число обслуживающих каналов, очередь отсутствует. В нем обрабатываются заявки клиентов на этапе установления соединения при осуществлении так называемого трехэтапного рукопожатия (three-way-handshake). Длительность обслуживания заявки каналом в данном центре равна времени «оборота» TCP сегмента RTT (Round Trip Time), т.е. времени прохождения его от сервера к клиенту и времени получения подтверждения на этот сегмент. Если в момент поступления сообщения в центр все K каналов заняты, то сообщение теряется, вероятность это события равна P_B. Время пребывания сообщений в данном центре ограниченно допустимым временем установления соединения и, при превышении данного времени, сообщения теряются с вероятностью P_{TO}. Дисциплина обслуживания в центре IS.

S2 – основной поток приложения сервера, извлекающего сообщения из очереди на установление соединения. Максимальная длина очереди L к центру задается в серверном приложении и, если при поступлении сообщения все L мест очереди заняты, то она теряется с вероятностью P_L. Дисциплина обслуживания в центре FCFS.

S3 – параллельные потоки сервера, обеспечивающие одновременное обслуживание соединений на этапе получения запросов по сети. Соответствует процессу передачи сообщения клиентом, а также передачи сегмента с подтверждением о получении сообщения. На данном этапе сообщение передается модулю аутентификации для обработки. Дисциплина обслуживания в центре IS.

Центры S3 и S8 имеют по M каналов обслуживания (потоков сервера) и при начале обслуживания сообщения в i-ом канале центра S3 он считается занятым до завершения обслуживания в i-ом канале центра S8. Таким образом, происходит блокировка каналов центров S3 и S8 и поэтому потерь сообщений из-за переполнения очереди к центрам S5, S6, S7 и занятости всех обслуживающих устройств центра S4 не происходит, т.к. больше чем M сообщений в центрах S4, S5, S6, S7 быть не может.

Времена пребывания сообщений в центрах S3 и S8 складываются из суммы времен передачи запроса от клиента и ответа сервера по каналам связи соответственно, и ожидания подтверждений на переданные данные. Минимальное время обработки сообщения (и запрос, и ответ помещаются в один стандартный TCP сегмент) в центре S3 равно RTT, в центре S8 – половине RTT. В общем случае, если запрос не удается передать в одном TCP сегменте (стандартно вмещает 1460 байт полезной информации), необходимо дополнительно учесть время на отправку полного запроса клиента серверу (центр S3) и время отправки ответа сервера клиенту (центр S8).

Если сообщение, обслуживание которой завершилось в центре S2, застает все M каналов центра S3 занятыми, оно блокируется и ожидает освобождения канала обслуживания в S3 в течение допустимого времени ожидания, при превышении которого сообщение теряется. Таким образом, время пребывания сообщений в центре S2 складывается из времени обслуживания, времени блокировки и времени ожидания в очереди. В случае возникновения ошибки в формате данных запросов с вероятностью P_{data} после обработки в центре S3, сообщения попадают в центр S8, где формируются ответы сервера, содержащие

данные об ошибке.

S4 – модуль аутентификации клиентов при обращении к хранилищу ЭД. Детали процесса аутентификации были рассмотрены ранее. Время пребывания сообщения в центре равно времени прохождения аутентификации пользователя по протоколу TLS. В случае если аутентификация была неудачна (вследствие недействительных сертификатов, ошибок при работе протокола TLS) клиенту отправляется ответ с отказом в доступе к КХЭД. Вероятность этого события – P_{auth}. Время пребывания сообщения в центре рассчитывается по формуле

$$10*\text{RTT} + 2* T_{OCSP} + 48/c\_asym + 500/c\_sym + T_s\,, \qquad (2.1)$$

где

- T_{OCSP} – время проверки статуса сертификата по протоколу OCSP;
- c_sym – скорость симметричного шифрования и дешифрования;
- c_asym – скорость асимметричного шифрования и дешифрования;
- T_s – время простановки ЭЦП.

Дисциплина обслуживания в центре IS.

S5 – модуль, формализующий процесс проверки прав доступа клиентов при обращении к хранилищу ЭД. Логика работы данного модуля требует выполнения криптографических операций, связанных с шифрованием и ЭЦП. В случае если клиент не обладает достаточными правами для работы с хранилищем, ему отправляется ответ с отказом в доступе. Вероятность этого события – P_{den}. Время пребывания сообщения в центре рассчитывается по формуле

$$(100/c\_asym + T_v) + (100/c\_asym + T_s) + (100/c\_sym + + T_v) + (100/c\_sym + T_s) + 0{,}15 = \qquad (2.2)$$

$$= 200/c\_asym + 200000/c\_sym + 2\,T_v + 2T_s + 0{,}15 \text{ с}$$

При прохождении процедуры проверки прав пользователей будем принимать во внимание следующие параметры:

- c_asym – скорость асимметричного шифрования и дешифрования;
- c_sym – скорость симметричного шифрования и дешифрования;
- T_s – время простановки ЭЦП.

- T_v – время проверки ЭЦП;
- T_{add} – время проведения вспомогательных операций.

Дисциплина обслуживания в центре FCFS.

В случае удачной аутентификации и проверки прав доступа клиента производится поиск ЭД по запросу пользователей и выполнение операций по контролю целостности информации, проверке и простановки ЭЦП, шифрованию и дешифрованию. Для формализации процесса индексного поиска и полнотекстового поиска отдельно выделены центры S6 и S7 соответственно. В связи с необходимостью выполнения трудоемких криптографических операций в процессе поиска, эти процессы формализуются в виде однолинейных центров с дисциплиной обслуживания FCFS, и длинами очередей равной M, где M – количество каналов обслуживания в центрах S3 и S8, т.е. максимальное количество потоков сервера приложений, обеспечивающих одновременное обслуживание соединений с пользователями. Время обслуживания в данных центрах рассчитывается по формулам:

$$T_{search} + T_v + T_v + Size_{doc} / c\_sym + T_v + T_s + T_{TSP} + T_{OCSP} + \\ + Size_{doc}/c\_sym + T_s + T_s + P_{get}*Size_{doc}/c\_sym = \\ = T_{search} + 3T_v + 3T_s + (2+P_{get})*Size_{doc}/c\_sym + T_{TSP} + T_{OCSP}. \quad (2.3)$$

Будем принимать во внимание следующие параметры:

- T_{search} – время индексного поиска электронного документа;
- T_{OCSP} – время проверки статуса сертификата по протоколу OCSP;
- T_{TSP} – длительность простановки метки времени;
- c_asym – скорость асимметричного шифрования и дешифрования;
- c_sym – скорость симметричного шифрования и дешифрования;
- T_s – время простановки ЭЦП;
- T_v – время проверки ЭЦП;
- $Size_{doc}$ – размер электронного документа.

$$(T_{search\_doc} + T_v + Size_{doc} / c\_sym)\, N_{doc}. \quad (2.4)$$

Будем принимать во внимание следующие параметры:

- T_{search_doc}– время поиска в документе;
- c_sym – скорость симметричного шифрования и дешифрования;
- T_v – время проверки ЭЦП;
- N_{doc} – количество ЭД, в которых осуществляется полнотекстовый поиск;
- $Size_{doc}$ – размер электронного документа.

Полнотекстовый поиск производится с вероятностью P_{txt}.

После обслуживания запроса пользователя в центрах S6 и S7 возможен повторный запрос на доступ к ЭД без прохождения аутентификации, в связи с чем предусмотрен переход к центру S5 с вероятностью P_{back}.

После того, как запрос пользователя выполнен, происходит передача ответа пользователю в многолинейном центре обслуживания S8. Дополнительно к ответу может прикрепляться найденный документ (в случае удачного поиска), вероятность этого события – $(1- P_{data})(1- P_{auth})(1- P_{den}) P_{get}$. Время обслуживания в данном центре принимается равным времени передачи ответа пользователю по каналу связи, защищенному протоколом TLS. Дисциплина обслуживания в центре IS.

Для рассматриваемой сети ненулевые вероятности переходов заявок между центрами (элементы матрицы переходных вероятностей) имеют следующий вид:

P[Si→S1]=1-P_B, P[S1→S2]= (1-P_{TO})(1-P_L),

P[S3→S4]= 1- P_{data}, P[S3→S8]= P_{data},

P[S4→S5]=1- P_{auth}, P[S4→S8]=P_{auth},

P[S5→S6]=1- P_{den}, P[S5→S8]=P_{den},

P[S6→S7]= P_{txt}, P[S6→S5]=(1- P_{txt}) P_{back}, P[S6→S8]=1- P_{txt},

P[S7→S8]=1- P_{back}, P[S7→S5]= P_{back},

P[S8→Ss]=1,

P[Si→Ss]= P_B, P[S1→Ss]= 1-(1-P_{TO})(1-P_L),

где Si - источник, а Ss - сток СеМО.

Для исследования разработанной модели СеМО аналитическими методами необходимо ввести ряд допущений:

1) входящий поток заявок должен быть пуассоновским;

2) каждый центр сети может быть представлен одним из четырех типов СМО, указанных в разделе 1.2 (для метода MVA) или центром типа FCFS (для метода Бузена);

3) распределение длительности обслуживания сообщений в центрах сети является экспоненциальным в центрах с дисциплиной обслуживания FCFS, либо общего вида для центров с дисциплинами IS, PS и LCFC;

4) длины очередей в центрах сети не ограниченны;

5) количество классов сообщений равно 1;

6) количество обслуживающих приборов в многолинейных центрах не ограниченно.

Необходимо отметить, что использование имитационного моделирования позволяет снять все эти ограничения.

Для учета особенностей работы описанных аналитических методов расчета, разомкнутую СеМО, формализующую функционирование КХЭД преобразуем к замкнутой. В полученной замкнутой СеМО заявки извне не поступают и не покидают сеть; количество заявок, циркулирующих в них постоянно и равно количеству конечных пользователей N.

Кроме того, в связи с указанными выше допущениями в аналитической модели СеМО будут отсутствовать потери заявок, и невозможен учет эффекта блокировок центров обслуживания.

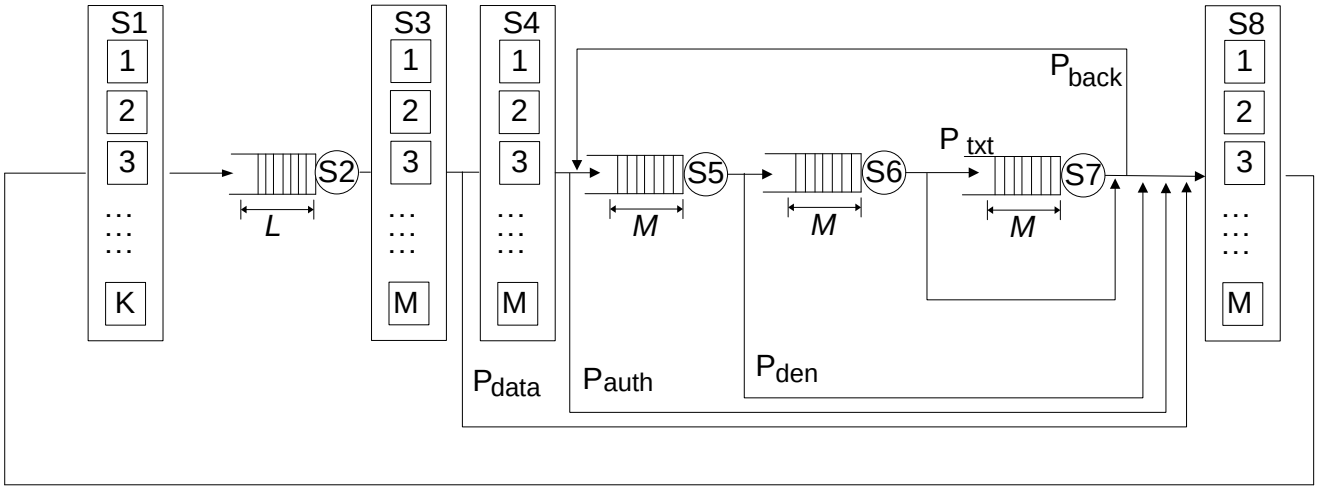

Рисунок 2.2. Структура замкнутой СеМО конфиденциального хранилища ЭД

В силу сделанных предположений структура замкнутой СеМО, формализующей функционирование конфиденциального хранилища ЭД будет выглядеть в соответствии с рисунок 2.2.

Обозначим через вектор $n=(n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8)$ состояние данной СеМО. Здесь n_i $(i=\overline{1,8})$ – число сообщений в i-ом центре. Через μ_i обозначим интенсивность обслуживания в i-м центре, тогда:

$$\mu_1 = n_1 / RTT,$$

$$\mu_2 = 1/T_a,$$

$$\mu_3 = \frac{n_3}{\frac{Size_{\mathrm{Re}q}}{b} + RTT + T_{route}} = \frac{bn_3}{Size_{\mathrm{Re}q} + bRTT + bT_{route}},$$

$$\mu_4 = n_4 / (10*\mathrm{RTT} + 2*T_{OCSP} + 48/c\_asym + 500/c\_sym + T_s),$$

$$\mu_5 = n_5 / (200/c\_asym + 200/c\_sym + 2\,T_v + 2T_s + T_{add}),$$

$$\mu_6 = n_6 / (T_{search} + 3T_v + 3T_s + (2+P_{get})*Size_{doc}/c\_sym + T_{TSP} + T_{OCSP}),$$

$$\mu_7 = n_7 / (T_{search} + 3T_v + 3T_s + (2+P_{get})*Size_{doc}/c\_sym + T_{TSP} + T_{OCSP} +$$

$$+ (T_{search\_doc} + T_v + Size_{doc}/c\_sym)\,N_{doc}),$$

$$\mu_8 = \frac{n_8}{\frac{Size_{Resp} + (1-P_{data})(1-P_{auth})(1-P_{den})\,P_{get}*Size_{doc}}{b} + \frac{RTT}{2} + T_{route}}.$$

где RTT – время “оборота” TCP пакета, T_a - время извлечения заявки из очереди на обслуживание и создания дочернего потока серверным приложением, T_{OCSP} – время проверки статуса сертификата по протоколу OCSP, *c_sym* – скорость симметричного шифрования и дешифрования, *c_asym* – скорость асимметричного шифрования и дешифрования, T_s – время простановки ЭЦП, T_v – время проверки ЭЦП, T_{search} – время индексного поиска электронного документа, T_{TSP} – длительность простановки метки времени, $Size_{doc}$ – размер электронного документа, N_{doc} – количество ЭД, в которых осуществляется полнотекстовый поиск, T_{search_doc}– время поиска в документе, $Size_{Req}$ – размер запроса клиента, $Size_{Resp}$ – размер ответа серверного приложения, b –скорость передачи данных в канале связи, защищенному протоколом TLS, T_{add} – время проведения вспомогательных операций..

Ненулевые вероятности переходов заявок между центрами следующие:

P[S1→S2]=1,

P[S2→S3]=1,

P[S3→S4]= 1- P_{data}, P[S3→S8]= P_{data},

P[S4→S5]=1- P_{auth}, P[S4→S8]= P_{auth},

P[S5→S6]=1- P_{den}, P[S5→S8]=P_{den},

P[S6→S7]= 1- P_{txt}, P[S6→S5]=(1- P_{txt}) P_{back}, P[S6→S8]=P_{txt},

P[S7→S8]=1- P_{back}, P[S7→S5]= P_{back},

P[S8→S1]=1,

Здесь S1, S3, S4 и S8 - IS центры, центры S2, S6, S7 - центр M/M/1 с дисциплиной обслуживания FCFS, S5 - M/G/1 с дисциплиной PS.

2.2 Расчет характеристик модели

Если принять за выделенный центр S1, то интенсивности потоков в центрах СеМО определяются следующим образом:

$$\lambda_i = e_i \Lambda, \quad i = \overline{1,R}$$

При e_1= 1 решением системы (1.7) для модели КХЭД будет:

$$e_1= e_2= e_3= 1, \tag{2.5}$$

$$e_4=1- P_{data}, \tag{2.6}$$

$$e_5= e_4P_{45} + e_6P_{65} + e_7P_{75} = \\ =(1- P_{data})(1- P_{auth}) + e_6(1- P_{txt}) P_{back} + e_7P_{back}, \tag{2.7}$$

$$e_6= e_5P_{56} = \\ =(1- P_{data})(1- P_{auth})(1- P_{den}) + e_6(1- P_{txt}) (1- P_{den}) P_{back} + \\ +e_7(1- P_{den})P_{back}, \tag{2.8}$$

$$e_7= e_6P_{67} = \\ =(1- P_{data})(1- P_{auth})(1- P_{den}) P_{txt} + \\ +e_6(1- P_{txt}) (1- P_{den}) P_{back} P_{txt}+e_7(1- P_{den})P_{back} P_{txt}, \tag{2.9}$$

$$e_8= 1. \tag{2.10}$$

Из (2.9) получаем:

$$e_7= [(1- P_{data})(1- P_{auth})(1- P_{den}) P_{txt} + \\ +e_6(1- P_{txt}) (1- P_{den}) P_{back} P_{txt}]/[1- (1- P_{den})P_{back} P_{txt}], \tag{2.11}$$

Подставляя (2.11) в (2.8), (2.7) и упрощая выражения получаем окончательно:

$$e_6= (1- P_{data})(1- P_{auth})(1- P_{den})/ [1- (1- P_{den})P_{back}] \\ e_7=(1- P_{data})(1- P_{auth})(1- P_{den}) P_{txt} /[1- (1- P_{den})P_{back}], \\ e_5=(1- P_{data})(1- P_{auth})/[1- (1- P_{den})P_{back}]. \tag{2.12}$$

Рассматриваемая замкнутая СеМО удовлетворяет требованиям теоремы ВСМР. В связи с тем, что количество классов сообщений в рассматриваемой сети равно 1, выражение для стационарной вероятности примет следующий вид ((1.3), (1.8), (1.9), (1.10)):

$$P(\bar{n}) = \frac{1}{G(N)}\left[\frac{1}{n_1!}\left(\frac{e_1}{\mu_1}\right)^{n_1}\right]\left[\frac{n_2!}{\mu_2^{n_2}}\left(\frac{e_2^{n_2}}{n_2}\right)\right]\left[\frac{1}{n_3!}\left(\frac{e_3}{\mu_3}\right)^{n_3}\right]\left[\frac{1}{n_4!}\left(\frac{e_4}{\mu_4}\right)^{n_4}\right]\left[\left(\frac{e_5}{\mu_5}\right)^{n_5}\right]\times$$

$$\times\left[\frac{n_6!}{\mu_6^{n_6}}\left(\frac{e_6^{n_6}}{n_6}\right)\right]\left[\frac{n_7!}{\mu_7^{n_7}}\left(\frac{e_7^{n_7}}{n_7}\right)\right]\left[\frac{1}{n_8!}\left(\frac{e_8}{\mu_8}\right)^{n_8}\right]=$$

$$=\frac{1}{G(N)}\frac{1}{n_1!}\left(\frac{RTT}{n_1}\right)^{n_1}\left(\frac{n_2!(T_a)^{n_2}}{n_2^{n_2+1}}\right)\frac{1}{n_3!}\left(\frac{\text{Size}_{\text{Req}}+bRTT+bT_{route}}{bn_3}\right)^{n_3}\times$$

$$\times\frac{1}{n_4!}\left(\frac{(1-P_{data})(10*RTT+2*T_{OCSP}+48/c\_asym+500/c\_sym+T_s)}{n_4}\right)^{n_4}\times$$

$$\times\left(\frac{(1-P_{data})(1-P_{auth})(200/c\_asym+200/c\_sym+2T_v+2T_s+T_{add})}{n_5}\right)^{n_5}\times$$

$$\times\frac{n_6!T_6^{n_6}((1-P_{data})(1-P_{auth})(1-P_{den})(1-P_{txt}))^{n_6}}{n_6^{n_6+1}}\times$$

$$\times\frac{n_7!T_7^{n_7}((1-P_{data})(1-P_{auth})(1-P_{den})P_{txt})^{n_7}}{n_7^{n_7+1}}\times$$

$$\times\frac{1}{n_8!}\left(\frac{2(Size_{Resp}+(1-P_{data})(1-P_{auth})(1-P_{den})P_{get}*Size_{doc})+bRTT+2bT_{route}}{2bn_8}\right)^{n_8}$$

где $T_6 = T_{search} + 3T_v + 3T_s + (2+P_{get})*Size_{doc}/c\_sym + T_{TSP} + T_{OCSP}$,

$T_7 = T_{search} + 3T_v + 3T_s + (2+P_{get})*Size_{doc}/c\_sym + T_{TSP} + T_{OCSP} +$

$+ (T_{search\_doc} + T_v + Size_{doc}/c\_sym)\, N_{doc})$.

Для вычисления основных характеристик СеМО (средних длин очередей и времен пребывания в центрах, пропускной способности и загрузки центров), будем использовать метод Бузена и метод анализа средних значений, описанные выше.

3. РАСЧЕТ ОТКРЫТОЙ СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ

Как было показано ранее, формализация КХЭД в виде замкнутой СеМО требует введения ряда существенных допущений, что значительно упрощает аналитическую модель. Формализация КХЭД в виде открытой сети массового обслуживания позволит отказаться от ряда допущений и получить результаты

более близкие к реальным характеристикам системы КХЭД, но потребует разработки методов расчета подобной аналитической модели. Далее будет предложен модифицированный метод анализа средних значений, который может быть использован для расчета открытой СеМО, формализующей систему КХЭД.

3.1. Модифицированный метода анализа средних значений

Поскольку рассматривается открытая СеМО, то возможны уходы заявок из сети, соответственно вычислять характеристики сети в зависимости от количества заявок в СеМО не представляется возможным. Будем производить все вычисления в зависимости от количества пользователей N, одновременно работающих с системой. На каждой итерации интенсивность входного потока сообщений $\lambda(N)$ будет рассчитываться как

$$\lambda(N) = \Lambda N, \qquad (3.1)$$

где Λ – интенсивность потока обращений к КХЭД каждого пользователя; N – количество пользователей.

Для формализации процесса поступления сообщений в СеМО, введем вспомогательный центр S0 (рисунок 3.1).

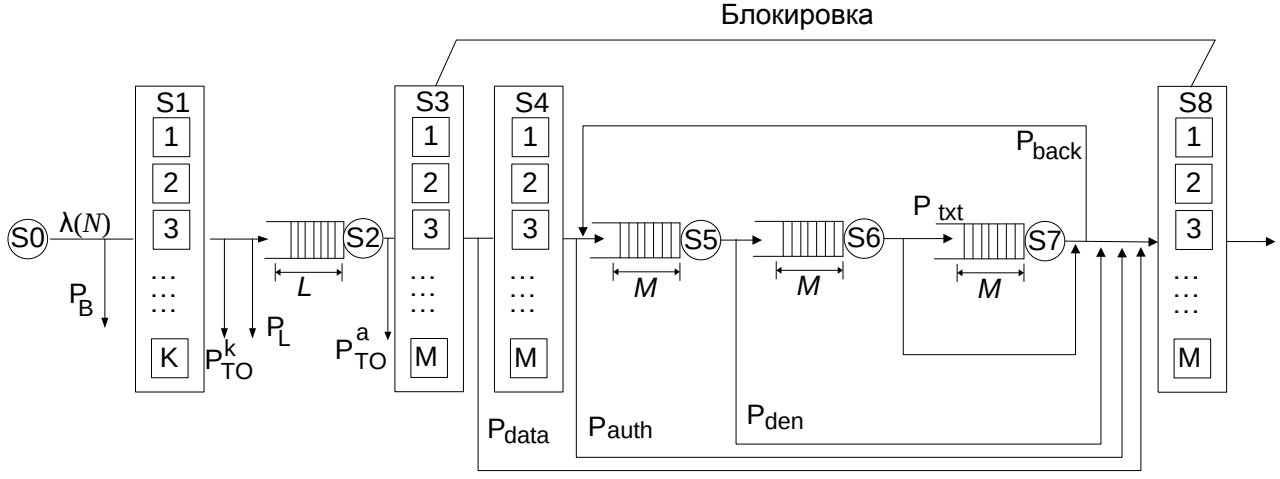

Рисунок 3.1. Открытая сеть массового обслуживания, формализующая работу КХЭД

Основная идея предлагаемого метода расчета характеристик открытой СеМО состоит в следующем:

1. Зная интенсивность входного потока сообщений в каждом центре, можно рассчитать интенсивность выходного потока и основные характеристики центра.

2. Поскольку число сообщений в СеМО не должно оставаться постоянным, можно учитывать потери из-за переполнения очередей ожидания и обслуживающих приборов центров.

3. Есть возможность учитывать эффект блокировки центров S3-S8. Для этого предварительно рассчитываются характеристики данных центров как для открытой СеМО. Это дает возможность определить суммарное число сообщений в центрах S3-S8 при заданной интенсивности входного потока сообщений. Далее необходимо учесть тот факт, что количество сообщений в центрах S3-S8 не может превышать определенного значения, и центры спроектированы так, чтобы вмещать любое количество сообщений в пределах заданного максимального значения. Соответственно, в стационарном режиме распределение сообщений в центрах S3-S8 фиксировано, и сообщения центры не покидают. Это дает возможность рассматривать центры S3-S8 как замкнутую подсеть массового обслуживания, и рассчитывать ее в соответствии с алгоритмом, приведенном в п. 1.1.2.

В предложенном подходе предоставляется возможность рассмотрения центров как связанных между собой элементов СеМО, что более полно отражает действительность и позволяет учесть потери и блокировки сообщений.

Рисунок 3.2 демонстрирует модифицированный алгоритм расчета открытой СеМО по методу MVA. В основе его лежит алгоритм, рассмотренный в п. 1.1.2, расчет производится итерационным методом по N (количество пользователей, одновременно работающих с КХЭД). В рамках одной итерации производятся следующие действия:

1. Рассчитывается интенсивность входного потока СеМО (равна

интенсивности выходного потока центра S0 – λ_0^{out}) в соответствии с (3.1) – шаг 1.

2. Рассчитывается время обработки сообщения для каждого центра – шаг 2.

3. Рассчитывается интенсивность входного потока для каждого центра λ_i^{in} – шаг 3, 4. В общем виде интенсивность входного потока равна сумме потоков, приходящих из других центров в соответствии с маршрутной матрицей.

$$\lambda_i^{in}(k) = \sum_{j=0}^{R} P_{ji}\lambda_j^{out}(k),\ i = \overline{1,R}, \qquad (3.2)$$

где P_{ij} – элементы маршрутной матрицы.

Однако произвести такой расчет не всегда возможно из-за наличия обратных связей – например от центра S7 к центру S5 (рисунок 3.1). В этом случае приходится рассматривать данный центр как независимый и определять интенсивность входного потока сообщений в соответствии с (1.5).

4. Определяется количество занятых обслуживающих приборов в центре – шаг 5.

5. Если рассчитанное количество занятых приборов превышает количество приборов центра, то интенсивность выходящего потока определяется интенсивностью обслуживания сообщения и количеством приборов центра – шаг 7, 8. Не попавшие на обслуживание сообщения либо помещаются в очередь, либо теряются (если очереди нет или она переполнена). Иначе, интенсивность выходного потока равна интенсивности входного – шаг 6.

6. Рассчитываем общее количество сообщений, находящихся в центре из соотношения (1.27) – шаг 9.

7. Если рассчитанное количество сообщений в центре превышает суммарный размер очереди и количество обслуживающих приборов, не попавшие на обслуживание и в очередь сообщения теряются – шаг 10.

8. Далее можно определить среднюю длину очереди L_i (k) – шаг 12.

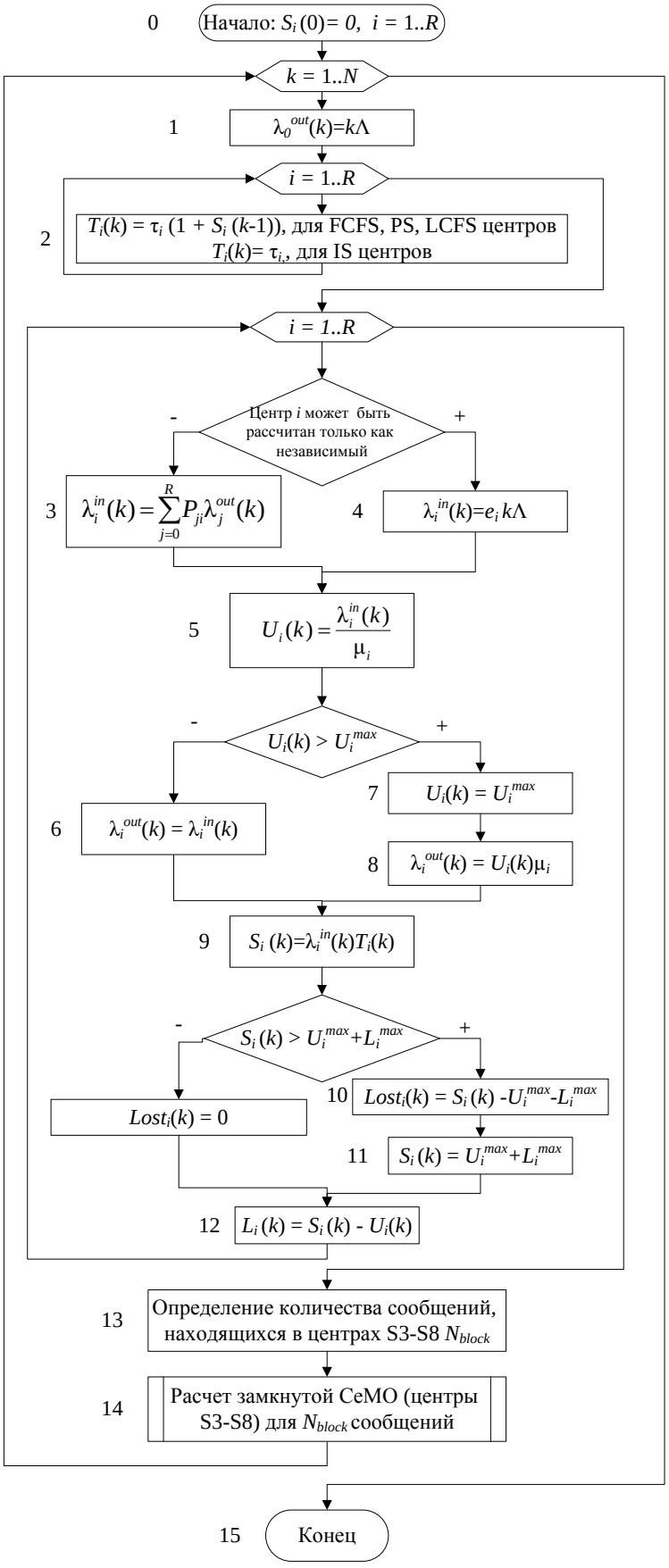

Рисунок 3.2. Алгоритм расчета открытой СеМО по методу анализа средних значений

10. После проведенных расчетов необходимо учесть эффект блокировки центров S3-S8. Поскольку сообщения в этих центрах теряться не могут (все потери должны происходить в центра S1, S2) можно формализовать обработку сообщений в этих центрах в виде функционирования замкнутой СеМО, находящейся в стационарном режиме. Режим открытой СеМО в данном случае не подходит, поскольку подразумевает постоянный входящий в сеть поток сообщений, чего не может быть при блокировке. Но расчет открытой СеМО необходим для определения количества сообщений, находящихся в центрах S3-S8, в зависимости от интенсивности потока сообщений от центра S0 – шаг 13. Расчет замкнутой сети для центров S3-S8 выполняется на шаге 14 по алгоритму, приведенному в п. 1.1.2.

4. ПОСТАНОВКА ЗАДАЧИ

Создать сеть массового обслуживания из 6-8 центров и рассчитать по одному из описанных выше методов.

5. ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ ДЛЯ ОПТИМИЗАЦИИ ПРОЦЕССА ВЫБОРА ПАРАМЕТРОВ КОНФИДЕНЦИАЛЬНОГО ХРАНИЛИЩА ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Имитационное моделирование является мощным аппаратом построения соответствующих моделей сложных систем. Оно обеспечивает глубокое представление моделируемого объекта, позволяет получить характеристики исследуемой системы на любом временном интервале, дает возможность учитывать случайные и неопределенные факторы. Каждая из характеристик сложной системы прозрачна для анализа при постоянных значениях прочих параметров. Но при рассмотрении динамики изменений всех параметров одновременно с целью их оптимизации требуется реализация серии

экспериментов на имитационной модели в большой области поиска. В действительно сложных системах пространство поиска может оказаться столь большим, что полный перебор будет является неприемлемым. Таким образом, встает задача поиска оптимальных параметров исследуемой системы методом, позволяющим значительно ускорить этот процесс на многомерном пространстве параметров по сравнению с полным перебором. В настоящее время для решения данной задачи широко применяются системы, основанные на генетических алгоритмах (ГА).

5.1. Простой генетический алгоритм

Рассмотрим простой генетический алгоритм, который может быть использован в процессе решения задачи функциональной оптимизации.

ГА состоит из следующих компонент:

1. Ресурсы:
   - **ген.** Наименьший элемент информации об особи (в технических задачах оптимизации соответствует 1 биту данных)
   - **хромосома**. Параметр или набор параметров для решаемой задачи. Состоит из генов.
   - **особь.** Решение рассматриваемой проблемы. Состоит из одной или нескольких хромосом.
   - **популяция**. Набор особей, представляющих различные решения задачи, за счет различных значений хромосом.
2. **Целевая функция** для оценки приспособленности (fitness) решений. Также носит название функции пригодности – критерий, рассчитываемый по определенным правилам.
3. Набор операторов, которые применяются к особям популяции для получения новых особей (решений проблемы):
   - оператор репродукции (reproduction).
   - оператор скрещивание (crossover).

- оператор мутация (mutation).

Оператор репродукции применяется к особям текущей популяции для отбора наиболее жизнеспособных и определения количества потомков, которые они могут дать. Основные виды репродукции:

- на основе рулетки. Отбираются особи с помощью n «запусков» рулетки. Колесо рулетки содержит по одному сектору для каждого члена популяции. Размер i-ого сектора пропорционален соответствующей величине $P_{sel}(i)$ вычисляемой по формуле:

$$P_{sel}(i) = \frac{f(i)}{\sum_{i=1}^{n} f(i)}, \tag{5.1}$$

где n – количество особей популяции; $f(i)$ – значение целевой функции i-й особи;

Для определения количества потомков (копий) n_i, которые может дать i-я особь для следующего поколения, вычисляют:

$$n_i = P_{sel}(i)n = \frac{f(i)n}{\sum_{i=1}^{n} f(i)}, \tag{5.2}$$

- на основе заданной шкалы. Популяция предварительно сортируется от «лучшей» особи к «худшей» на основе заданного критерия. Каждому элементу назначается определенное число и далее репродукция выполняется согласно этому числу.
- элитная репродукция. В этом случае выбираются лучшие (элитные) особи на основе сравнения целевой функции. Далее они вступают в различные преобразования, после которых снова выбираются элитные элементы. Процесс идет до тех пор, пока продолжают появляться элитные особи.
- турнирная репродукция. Реализует n турниров, чтобы выбрать n особей. Каждый турнир построен на выборке k элементов из популяции, и выбора лучшей особи среди них. Наиболее распространен турнирный отбор с k=2.

В данной работе будет применяться оператор репродукции на основе рулетки.

Оператор скрещивание (crossover) осуществляет обмен частями хромосом особи между двумя (может быть и больше) хромосомами особей в популяции. Наиболее простым является одноточечное скрещивание, работающее следующим образом: сначала случайным образом выбирается точка разрыва. Точка разрыва – участок между соседними битами в строке. Обе родительские структуры разрываются на два сегмента по этой точке. Затем родители обмениваются сегментами, находящимися до или после точки разрыва, между собой и получаются два генотипа потомков (рисунок 5.1).

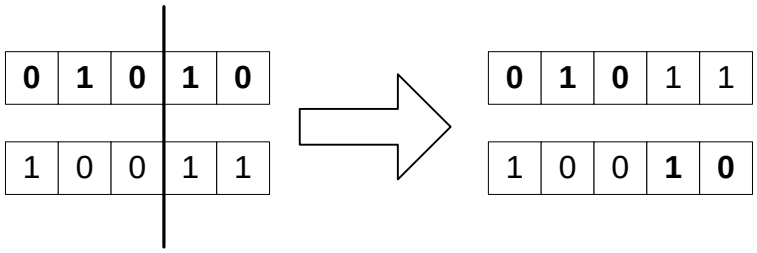

Рисунок 5.1. Скрещивание особей

Оператор мутации (mutation) инициирует стохастическое изменение части хромосом. Мутация необходимо потому, что предотвращает потерю важного генетического материала. Как правило, применяют одноточечный оператор мутации, который случайно выбирает ген в хромосоме и обменивает его на рядом расположенный (рисунок 5.2).

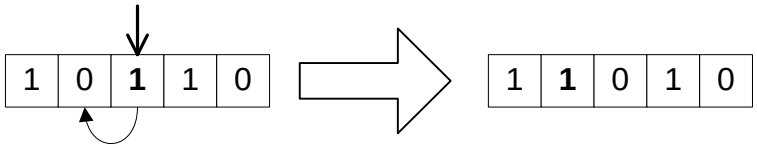

Рисунок 5.2. Мутация

Работа ГА представляет собой итерационный процесс, который продолжается до тех пор, пока не будет воспроизведено заданное число поколений или какой-либо иной критерий останова. На каждом поколении ГА

реализуется репродукцию пропорционально приспособленности, скрещивание и мутация.

Хотя модель эволюционного развития, применяемая в ГА, сильно упрощена по сравнению со своим природным аналогом, тем не менее ГА является достаточно мощным средством и может с успехом применяться для широкого класса прикладных задач, включая те, которые трудно, а иногда и вовсе невозможно, решить другими методам. Однако, ГА, как и другие методы эволюционных вычислений, не гарантирует обнаружения глобального решения за полиномиальное время. Главным же преимуществом генетических алгоритмов является то, что они могут применяться даже на сложных задачах, там, где не существует никаких специальных методов. Даже там, где хорошо работаю существующие методики, можно достигнуть улучшения сочетанием их с ГА.

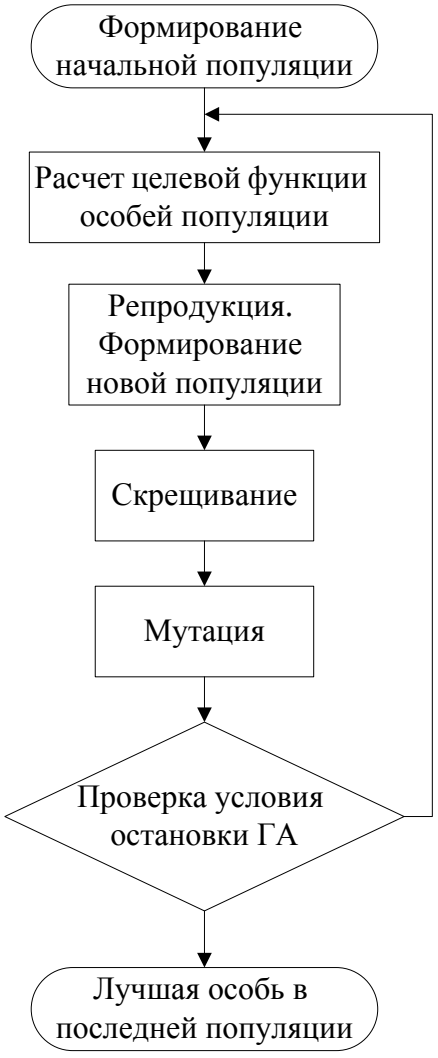

Рисунок 5.3. Простой генетический алгоритм

5.2 Гибридная система оптимизации параметров конфиденциального хранилища электронных документов

Разработка любой сложной технической системы требует прохождения определенного набора этапов, состоящего, в общем случае, из анализа, проектирования, эволюции, модификации. Причем этот процесс является циклическим. Выполнение любого этапа проводится с использованием имитационных моделей системы, которые обеспечивают глубокое представление моделируемого объекта, дают возможность анализа процессов на любом временном интервале, позволяют учитывать случайные и неопределенные факторы, оценивать как технические, так и экономические показатели функционирования системы. С другой стороны, наличие большого числа гетерогенных подсистем требует разработке комплексных решений по анализу и управлению такими системами, что обеспечит пользователя поддержкой при принятии решений.

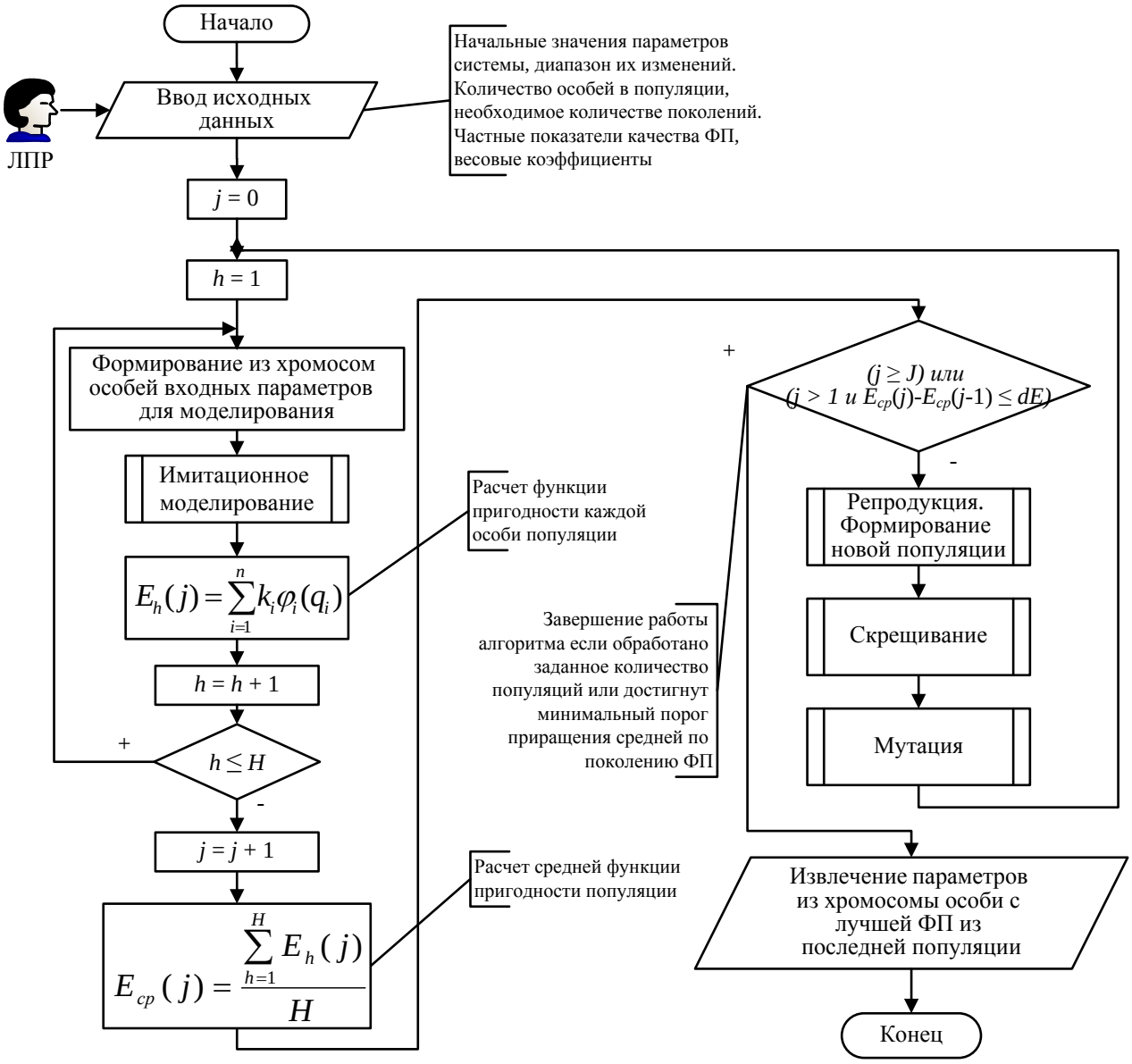

Рисунок 5.4. Гибридная система с генетическим алгоритмом и имитационной моделью

Существование ситуации, когда необходимо совместно использовать имитацию и различные методы принятия решений привело к появлению так называемых гибридных систем. Под гибридной системой в данной работе понимается система, объединяющая в себе имитационную модель и блок оптимизации. Блок оптимизации реализует ГА, а имитационная модель служит для вычисления значения критерия оптимизации (являющегося целевой функцией или функцией пригодности) для выбираемых вариантов решения (рисунок 5.4). На рисунке h – номер особи в популяции, H – количество особей в каждой

популяции, j – номер популяции, J – количество популяций, dE – значение порога минимального приращения функции пригодности.

К преимуществам использования ГА в системах оптимизации можно отнести возможность использования нескольких начальных альтернативных решений, возможность одновременного изменения нескольких параметров в процессе поиска оптимального решения, возможность «выхода» из локальных оптимумов.

В гибридных системах целевая функция имеет название функции пригодности и рассчитывается в соответствии с требованиями ЛПР. Входными данными для имитационного моделирования являются параметры, содержащиеся в хромосомах особей-решений. Таким образом, получив по итогам работы гибридной системы особь с лучшей ФП, можно определить параметры исходной системы, которые соответствуют этому оптимальному решению.

Далее рассмотрим гибридную систему оптимизации с имитационной моделью КХЭД в качестве решающего блока и простейшим ГА в качестве оптимизационного блока. Такая система позволит находить оптимальные параметры функционирования конфиденциального хранилища ЭД с учетом ограничений, накладываемых на диапазон изменений важнейших параметров и оптимальности получаемого результата моделирования. Параметры, способные изменять свои значения в заданном диапазоне и вид функции пригодности, используемой для оценки оптимальности, определяет ЛПР.

Приведем пример исследования и расчета оптимальных параметров КХЭД. Опираясь на результаты аналитического и имитационного моделирования, выделим основные параметры, подлежащие оптимизации:

- M – максимальное количество используемых серверным приложением параллельных потоков, обеспечивающих обслуживание заявок пользователей;
- L – максимальная длина очереди на обслуживание;
- T_{OCSP} – время проверки статуса сертификата по протоколу OCSP;

- T_{search} – время индексного поиска электронного документа;
- T^{s}_{out} – допустимое время ожидания начала обслуживания;
- *c_sym* – скорость симметричного шифрования и дешифрования;
- *b* – скорость передачи данных в канале связи, защищенном по протоколу TLS;

Отдельно стоит отметить важность параметров, характеризующих нагрузку на систему:

- *N* – число пользователей системы;
- *Labmda* – интенсивность генерации новых сообщений каждым пользователем;

Для определения этих параметров должен проводиться анализ существующего документооборота, определяется состав и количество пользователей, участвующих в активной работе с электронными документами. Также должна учитываться возможность масштабирования системы.

Кроме того, будем учитывать важнейшие выходные характеристики системы:

- T_{cp} – среднее время обработки одного сообщения в системе;
- P^{q}_{lost} – коэффициент потерь заявок из-за переполнения очереди сервера приложений;
- P^{t}_{lost} – коэффициент потерь заявок из-за превышения допустимого времени ожидания в очереди сервера приложений;
- *M* – количество используемых серверным приложением параллельных потоков;
- *W* – пропускная способность системы.

Параметрами простого генетического алгоритма являются:

- *Generation_Size* – размер популяции (количество особей в популяции);
- *Generations_Count* – количество воспроизводимых популяций;
- P_{cross} – вероятность скрещивания (определяет, с какой вероятностью произойдет скрещивание данной пары особей);

- P_{mut} – вероятность мутации (определяет, с какой вероятностью произойдет мутация данной особи).

Для оптимизационных задач, вероятность скрещивания обычно принимают равной 0,6 – 0,99, а вероятность мутации от 0,6 и меньше.

Выбор размера популяции, при котором алгоритм показывает наилучшие результаты, является, в общем случае, нетривиальной задачей, поскольку комплексно зависит от других параметров генетического алгоритма, а также от поставленной проблемы. Выбор недостаточно большого числа особей в популяции может привести к преждевременной сходимости решений вокруг локального оптимума. Кроме того, возможно появление таких областей, которые не могут быть исследованы оператором скрещивания, что отрицательно влияет на надежность алгоритма (в такой области может оказаться глобальный экстремум). Задание большого размера популяции приведет к неоправданно большому времени работы алгоритма.

В стационарных ГА с постоянным количеством особей в поколении авторы предлагают подход, при котором размер популяции задает ЛПР. В то же время имеются работы, в которых предлагается следующий подход.

Введем следующие обозначения:

- N – размер популяции;
- L – размер хромосомы;
- b^j_i – i-й бит j-й хромосомы;

S^N_L – событие, когда после инициализации ГА с размером популяции N и числом бит в хромосоме L для любого $0 \le k \le L-1$ выполняется неравенство:

$$0 < \sum_{j=1}^{N} b_k^j < N, \qquad (5.3)$$

т.е. событие S^N_L фиксирует тот факт, что нет ни одной позиции в битовой строке хромосом, для которой биты в этой позиции у всех особей начальной популяции были бы одинаковыми.

Очевидно, что для двух однобитных хромосом

$$P(\overline{S_1^2}) = 2 \cdot \left(\frac{1}{2}\right)^2. \qquad (5.4)$$

Поскольку при инициализации каждый бит хромосомы инициализируется независимо от других битов, то в общем случае вероятность наступления события S_L^N будет определяться следующим выражением:

$$P(S_L^N) = \left(1 - 2 \cdot \left(\tfrac{1}{2}\right)^N\right)^L, \qquad (5.5)$$

Откуда

$$N = \left\lfloor \log_{0.5}\left(1 - 2 \cdot \left(\frac{1}{2}\right)^N\right)^L \right\rfloor. \qquad (5.6)$$

Вероятность $P(S_L^N)$ выбирается близкой к 1. Так, для хромосомы длиной *L* = 40 бит и $P(S_L^N)$=0.999 целесообразно задавать *N* не менее 17 особей.

5.3 Оценка функции пригодности КХЭД

Поскольку необходимо провести оценку нескольких характеристик КХЭД, функцию пригодности можно представить в форме обобщенного критерия, который связывает эффективность системы со всеми частными показателями. Например, одним из наиболее распространенных форм обобщенного критерия является нормированный аддитивный критерий:

$$E = \sum_{i=1}^{n} k_i \varphi_i(q_i), \qquad (5.7)$$

где q_i – частный показатель качества (характеристика) системы КХЭД; n – количество анализируемых характеристик; функция $\varphi_i(q_i)$ подбирается так, чтобы исключить размерность *i*-й характеристики и обеспечить условие $\varphi_i(q_i) \in [0, 1]$, а весовые коэффициенты k_i удовлетворяют условию

$$\sum_{i=1}^{n} k_i = 1, k_i > 0. \qquad (5.8)$$

Но при этом возникает проблема определений значений весовых коэффициентов, которая может быть решена различными методами. В данной

работе весовые коэффициенты определяет ЛПР. В случае использования аддитивного критерия ФП будет иметь следующий вид:

$$E = k_1\varphi_i(T_{cp}) + k_2\varphi_i(P^q_{lost}) + k_3\varphi_i(P^t_{lost}) + k_3\varphi_i(M) + k_3\varphi_i(W). \quad (5.9)$$

Выбор данного вида функции пригодности объясняется следующими причинами: присутствие частных показателей качества различных размерностей требует их нормировки к единому значению по определенному закону, кроме того значение или полезность каждого из частных показателей может быть различным в каждом конкретном случае, что потребовало введения весовых коэффициентов. Аддитивный вид результирующей ФП позволяет учесть ситуации, в которых высокие значения отдельных частных показателей компенсируют низкие значения других показателей, что может соответствовать пожеланиям ЛПР. К примеру, быстрота работы системы в целом может быть гораздо важнее затрат на приобретение дополнительной оперативной памяти. И если удастся подобрать параметры системы в соответствии с данными условиями, результирующая ФП будет иметь высокие значения, что будет вполне соответствовать пожеланиям ЛПР. С другой стороны, применение, например, мультипликативной ФП в таком случае дало низкие значения результирующего показателя эффективности, и это шло бы в разрез с реальной ситуацией.

Далее эмпирически подберем частные нормированные ФП. ФП среднего времени обработки заявки наиболее адекватно описывается функцией вероятности отказа систем с избыточностью, которая имеет следующий вид (рисунок 5.5):

$$p = 1 - (1 - e^{-\lambda t})^m. \quad (5.10)$$

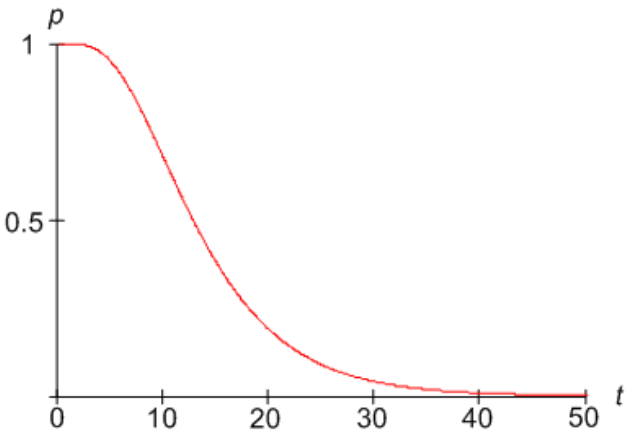

Рисунок 5.5. Вид частной функции пригодности среднего времени обработки сообщения

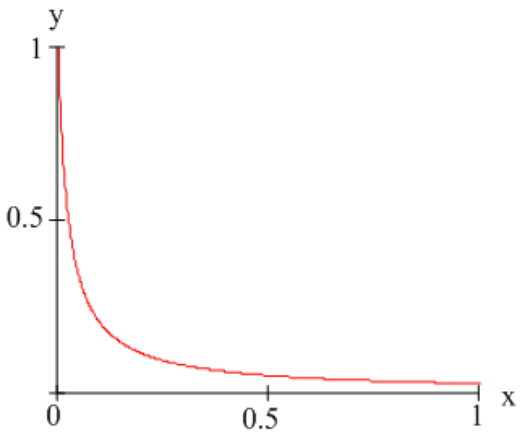

Рисунок 5.6. Вид частной функции пригодности коэффициентов потерь сообщений

При этом параметры λ и m в контексте применения в определении пригодности соответственно равны 0,2 и 15 (значения могут быть изменены ЛПР с учетом конкретной ситуации).

ФП коэффициентов потерь можно представить функциями вида

$$y = \frac{a}{b + cx}. \qquad (5.11)$$

Действительно, появление необслуженных сообщений в КХЭД приведет к нарушению нормальной работы с документами и, как следствие, нарушению процесса документооборота и потере данных. Поэтому, даже небольшое увеличение коэффициентов потерь (относительно нуля) должно вести к резкому снижению ФП. На графике (рисунок 5.6) показана частная ФП для коэффициентов потерь: a = b = 0.5, c = 20.

Параметр M определяет количество потоков сервера, которое должно оцениваться как выходными характеристиками системы, так и стоимостью системы. Для учета затрат на повышение объема ОЗУ воспользуемся линейной зависимостью (рисунок 5.7). Данная функция показывает обратную зависимость полезности критерия от стоимости:

$$y = 1 - 0{,}002M. \qquad (5.12)$$

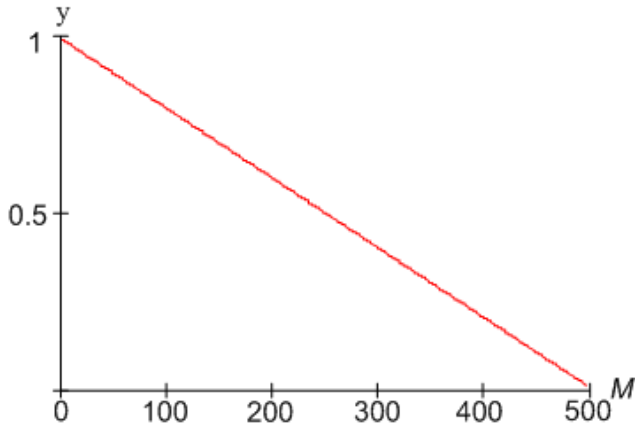

Рисунок 5.7. Вид частной функции пригодности количества потоков сервере

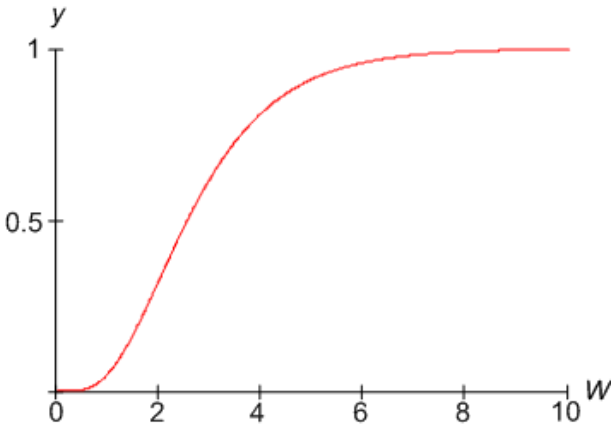

Рисунок 5.8. Вид частной функции пригодности пропускной способности системы

Пропускная способность системы – характеристика, увеличение которой говорит о повышении эффективности работы системы. В данном случае выберем для ее описания следующую зависимость (рисунок 5.8). Данный вид графика говорит о том, что значение пропускной способности, начиная с определенного значения, принимается как вполне удовлетворяющее ЛПР:

$$y = (1- e^{-\lambda W})^m. \qquad (5.13)$$

При этом параметры λ и m в контексте применения в определении пригодности соответственно равны 1 и 10.

Вид функций пригодности для каждой из частных характеристик должен выбираться ЛПР. Условием остановки работы ГА может быть завершение обработки некоторого количества поколений или достижение минимального порога приращения средней по поколению ФП после обработки очередной популяции.

5.4 Результаты работы гибридной системы оптимизации

Гибридная система оптимизации была реализована в среде моделирования AnyLogic 6.4.1 включает в себя имитационную модель КХЭД и описанный ранее простой генетический алгоритм.

Таблица 5.1 Функции пригодности особей, полученные в результате работы гибридной системы оптимизации параметров КХЭД

Популяция / Особь	1	10	20	30	40
1	0.5111	0.9349	0.9545	0.9212	0.8744
2	0.9140	0.6884	0.9412	0.9143	0.9148
3	0.5759	0.6491	0.9145	0.9347	0.9095
4	0.5182	0.9272	0.9148	0.9036	0.9148
5	0.4894	0.9140	0.5194	0.9267	**0.9649**
6	0.5568	0.4901	0.9148	0.8872	0.9349
7	0.6672	0.5448	0.9001	0.9085	0.8560
8	0.5249	0.6520	0.9112	0.9184	0.9140
9	0.6793	0.9140	0.9112	0.9331	0.9146
10	0.8070	0.9148	0.9159	0.8892	0.9148
11	0.4986	0.5153	0.9154	0.9127	0.9148
12	0.7617	0.6820	0.8759	0.9152	0.9149
13	0.4760	0.9250	0.5043	0.9342	0.9211
14	0.4923	0.4688	0.7995	0.9101	0.9148
15	0.5276	0.6230	0.9151	0.9109	0.9149
16	0.4967	0.9148	0.8359	0.8884	0.9148
17	0.6360	0.8047	0.9203	0.8895	0.9587
18	0.5545	0.9148	0.9273	0.9147	0.9015
19	0.5206	0.9149	0.9141	0.8762	0.9149
20	0.7855	0.6700	0.9201	0.8892	0.9140

В эксперименте параметры ГА имели следующие значения:

- количество особей в популяции - 20;
- количество обрабатываемых популяций – 40;
- вероятность скрещивания – 0.7;
- вероятность мутации – 0.2;
- $k_i\,(T_{cp}, P^q_{lost}, P^t_{lost}, M, W)$=[0,7; 0,1; 0,1; 0,05; 0,05].

Параметры требования к КХЭД по нагрузке:

- количество пользователей – 1000;
- интенсивность генерации сообщений каждого пользователя 0.01 сообщений/с;

Таблица 5.2 Результаты работы гибридной системы оптимизации параметров КХЭД

Наименование параметра (характеристики)	Заданный диапазон изменений	Исходное значение	Оптимизи-рованное значение
Максимальное количество используемых серверным приложением параллельных потоков	[1, 500]	256	134
Максимальная длина очереди на обслуживание	[0, 2000]	20	48
Время проверки статуса сертификата по протоколу OCSP (с)	[0,02; 30]	2	0,9
Время индексного поиска электронного документа (с)	[0,05; 10]	5	0,5
Допустимое время ожидания начала обслуживания (с)	[0, 50]	3	14
Скорость симметричного шифрования и дешифрования (Кбайт/с)	[100, 70000]	66000	32400
Пропускная способность канала связи, защищенного по протоколу TLS (Кбайт/с)	[5, 250]	32	102
Среднее время обработки сообщения (с)		392	22,8
Коэффициент потерь сообщений из-за переполнения очереди ожидания		0,74	0,08
Коэффициент потерь сообщений из-за превышения времени ожидания		0,12	0,03
Пропускная способность системы (сообщ/с)		0,12	1,83

Область определения параметров функционирования системы:

- $M \in [1, 500]$;
- $L \in [0, 2000]$;
- $T_{OCSP} \in [0{,}02; 30]$ с;
- $T_{search} \in [0{,}05; 10]$ с;
- $T^{s}_{out} \in [0, 50]$ с;
- $c\_sym \in [100, 70000]$ Кбайт/с;
- $b \in [32, 250]$ Кбайт/с;

Результаты работы гибридной системы демонстрирует таблица 5.1. Как видно из результатов расчета лучшей ФП обладает особь №5 из 40-го поколения. Для нее были получены следующие значения параметров (Таблица 5.2):

Анализируя полученные значения, можно сказать, что наиболее критичным являются временные параметры индексного поиска ЭД и проверки статуса сертификатов, что требует от проектировщиков КХЭД увеличения производительности сервера баз данных и, возможно, создания корпоративного УЦ. Количество потоков сервера приложений почти вдвое меньше, чем закладывалось при имитационном моделировании, что позволит сэкономить на ОЗУ. Скорость шифрования также вдвое меньше запланированной, что говорит о не приоритетном влиянии данного параметра на производительность системы в целом (что соответствует результатам аналитического и имитационного моделирования). Пропускная способность каналов связи требует значительного увеличения, что должно быть учтено при проектировании КХЭД (модернизация каналов связи и коммутационного оборудования).

Как видно из примера, применение гибридной системы оптимизации позволило подобрать параметры системы, уменьшающие среднее время обработки сообщений более чем в 17 раз, суммарный коэффициент потерь сообщений более чем в 10 раз, увеличить пропускную способность системы более чем в 15 раз. Это дает возможность повысить эффективность работы системы, а полученные параметры позволяют обоснованно принимать такие решения как: внесение изменений в сетевую инфраструктуру (для увеличения пропускной способности каналов связи), уменьшение объема оперативной памяти сервера приложений (допустимо благодаря уменьшению количества потоков сервера), применение в КХЭД более быстрых механизмов индексного поиска и т.д.

5.5 Метод оптимизации параметров конфиденциального хранилища электронных документов

Анализ и проектирование конфиденциального хранилища электронных документов является сложной задачей, решение которой связано с проведением

многочисленных экспериментальных и аналитических исследований. Результатом исследования будут являться оценки показателей эффективности функционирования КХЭД или его компонент при заданных значениях параметров. В настоящей работе основное внимание уделено рассмотрению вопросов исследования и анализа процессов обработки и передачи информации в КХЭД и разработке аналитических и имитационных моделей СеМО, обеспечивающих комплексную оценку эффективности функционирования системы КХЭД, а также разработке гибридной системы. Поэтому предлагаемый метод оптимизации параметров функционирования КХЭД основывается на использовании разработанных моделях, алгоритмах и гибридной системе. Метод состоит из следующих этапов:

1. Определение требований, предъявляемых к КХЭД. Для этого проводится анализ существующего документооборота, определяется количество пользователей, участвующих в информационном обмене с использованием средств криптографической защиты информации, определяются виды электронных документов, их размер, анализируются возможности используемого программно-аппаратного обеспечения, сети передачи данных, коммутационного оборудования, необходимых для организации доступа к хранилищу ЭД конечных пользователей. На основании данного анализа формируются требования к КХЭД, в частности, по нагрузке на систему (количестве запросов в единицу времени, проценту запросов на проверку цифровых сертификатов), масштабируемости, стоимости, возможности модернизации, техническим характеристикам и т.д. Данные требования должны найти отражение в значениях входных параметров и допустимых диапазонах их изменения.

2. С учетом выбранных требований определяется набор показателей качества функционирования систем, которые необходимо оптимизировать.

3. Осуществляется сбор данных о параметрах функционирования системы и ее компонент. Собираются данные о рабочей нагрузке. Многие параметры – это случайные величины, поэтому для их определения организуется

сбор статистики и последующая ее обработка. Для аналитических моделей значения параметров принимаются детерминированными. При этом для получения значений характеристик с заданной точностью при имитационном моделировании определяется требуемое количество экспериментов.

4. ЛПР выбирает нормированные ФП для каждого из частных критериев оценки и определяет значения весовых коэффициентов для каждого критерия.

5. Устанавливаются диапазоны изменения характеристик системы и значений параметров функционирования.

6. Осуществляется поиск оптимальных параметров при помощи ПГА и имитационной модели.

7. После получения значений требуемых характеристик, производится их анализ, проверка адекватности и, в случае необходимости, корректировка параметров или самой модели.

Таким образом, приведенный метод позволяет осуществлять проектирование и исследовать существующие системы конфиденциальных хранилищ электронных документов с целью повышения эффективности и обоснования параметров функционирования КХЭД. Применение предложенного метода и разработанной гибридной системы позволяет выявлять ошибки при создании КХЭД еще на этапе проектирования, что может позволить сэкономить финансовые и временные затраты на дальнейшую модернизацию и доработку хранилища. Кроме того, использование разработанной гибридной системы позволяет обоснованно выбирать значения параметров технических компонент КХЭД уже при его эксплуатации в условиях меняющихся внешних и внутренних факторов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Benjamin Lee, «An Architectural Assessment of SPEC CPU Benchmark Relevance», Harvard University, Cambridge, MA, Tech.Rep. TR-02-06, 2006.

2. Control Objectives for Information and related Technology (COBIT). [Электронный ресурс]. – 2007. – Режим доступа: http://www.isaca.org/

3. D. Chandra, F. Guo, S. Kim, and Y. Solihin, “Predicting Inter-Thread Cache Contention on a Multi-Processor Architecture”, In Proc. Of 11th Int’l. Symposium on High-Performance Computer Architecture, pp. 340-351, 2005.

4. Decision Support System Consulting [Электронный ресурс]. – 2009. – Режим доступа: http://www.dssconsulting.ru/services/marketing/analytics/?id=51.

5. J. Hillston, Fluid flow approximation of PEPA models, in: Proceedings of QEST’05, pp. 33-43, IEEE Computer Society, 2005.

6. N. Thomas and Y. Zhao, Fluid flow analysis of a model of a secure key distribution centre, in: Proceedings 24th Annual UK Performance Engineering Workshop, Imperial College London, 2008.

7. Вишневский В.М. Теоретические основы проектирования компьютерных сетей. – М.: Техносфера, 2003. – 512с.

8. Ивницкий В.А., Теория сетей массового обслуживания. – М.:Физматлит, 2004. - 770 стр.

9. Саульев В.К. Математические методы теории массового обслуживания. - М.: Статистика, 1979.