Параметры и характеристики СМО

Параметры СМО

Для описания СМО используются три группы параметров:

- структурные;
- нагрузочные;
- функциональные параметры (параметры управления).

К *структурным параметрам* относятся:

- *количество обслуживающих приборов* K, равное 1 для одноканальной СМО и $K > 1$ для многоканальной СМО;
- *количество* k *и ёмкости накопителей* E_j $(j = \overline{1,k})$;
- *способ взаимосвязи накопителей с приборами* (в случае многоканальных СМО), например в виде матрицы связей.

Нагрузочные параметры СМО включают в себя:

- количество поступающих в систему классов заявок H, которое равно 1 для СМО с однородным потоком заявок и $H > 1$ для СМО с неоднородным потоком;
- закон распределения $A_i(\tau)$ интервалов времени между поступающими в систему заявками класса $i = \overline{1,H}$ или, по-крайней мере, первые два момента распределения, задаваемые, например, в виде интенсивности λ_i и коэффициента вариации ν_{a_i} интервалов;
- закон распределения $B_i(\tau)$ длительности обслуживания заявок класса $i = \overline{1,H}$ или, как минимум, первые два момента распределения, в качестве которых обычно используются средняя длительность b_i или интенсивность $\mu_i = 1/b_i$ обслуживания и коэффициент вариации ν_{b_i}.

Обозначения СМО (символика Кендалла)

Для компактного описания систем массового обслуживания часто используются обозначения, предложенные Д. Кендаллом [9], в виде:

A/B/N/L ,

где **A** и **B** – задают законы распределений соответственно интервалов времени между моментами поступления заявок в систему и длительности обслуживания заявок в приборе; **N** – число обслуживающих приборов в системе $(N = 1, 2, ..., \infty)$; **L** – число мест в накопителе, которое может принимать значения 0, 1, 2, … (отсутствие **L** означает, что накопитель имеет неограниченную ёмкость).

Для задания законов распределений **A** и **B** используются следующие обозначения:

G (General) – произвольное распределение общего вида;

M (Markovian) – экспоненциальное (показательное) распределение;

D (Deterministik) – детерминированное распределение;

U (Uniform) – равномерное распределение;

$\mathbf{E_k}$ (Erlangian) – распределение Эрланга *k*-го порядка (с *k* последовательными одинаковыми экспоненциальными фазами);

$\mathbf{h_k}$ (hipoexponential) – гипоэкспоненциальное распределение *k*-го порядка (с *k* последовательными разными экспоненциальными фазами);

$\mathbf{H_r}$ (Hiperexponential) – гиперэкпоненциальное распределение порядка *r* (с *r* параллельными экспоненциальными фазами);

g (gamma) – гамма-распределение;

P (Pareto) – распределение Парето и т.д.

Примеры:

M/M/1 – одноканальная СМО с накопителем неограниченной ёмкости, в которую поступает однородный поток заявок с экспоненциальным распределением интервалов времени между последовательными заявками (простейший поток) и экспоненциальной длительностью обслуживания заявок в приборе.

M/G/3/10 – трёхканальная СМО с накопителем ограниченной ёмкости, равной 10, в которую поступает однородный поток заявок с экспоненциальным распределением интервалов времени между последовательными заявками (простейший поток) и длительностью обслуживания заявок, распределённой по закону общего вида.

D/$\mathbf{E_2}$/7/**0** – семиканальная СМО без накопителя (ёмкость накопителя равна 0), в которую поступает однородный поток заявок с детерминированными интервалами времени между последовательными заявками (детерминированный поток) и длительностью обслуживания заявок в приборе, распределённой по закону Эрланга 2-го порядка.

Режимы функционирования СМО

СМО может работать в следующих режимах:

- **установившемся** или **стационарном**, когда вероятностные характеристики системы не изменяются со временем;
- **неустановившемся**, когда характеристики системы изменяются со временем, что может быть обусловлено:
  - *началом работы системы*, когда значения характеристик функционирования, меняясь со временем, стремятся в пределе к стационарным значениям (**переходной режим**);
  - *нестационарным характером* потока заявок и обслуживания в приборе (**нестационарный режим**).

Характеристики СМО с однородным потоком заявок

Характеристики систем со стохастическим характером функционирования являются *случайными величинами* и полностью описываются соответствующими законами распределений. На практике при моделировании часто ограничиваются определением только *средних значений* (математических ожиданий), реже – определением двух первых моментов этих характеристик.

В качестве основных характеристик СМО с однородным потоком заявок используются следующие величины:

- ***нагрузка*** системы:

$$\boxed{y = \lambda / \mu = \lambda b}; \qquad (3.6)$$

- ***коэффициент загрузки*** или просто ***загрузка*** системы, определяемая как доля времени, в течение которого система (в случае одноканальной СМО – прибор) работает, то есть выполняет обслуживание заявок; загрузка может быть рассчитана как отношение *среднего* времени T_p работы одного прибора многоканальной СМО, к общему времени наблюдения T:

$$\rho = \lim_{T \to \infty} \frac{T_p}{T}; \qquad (3.7)$$

время T_p для СМО с K обслуживающими приборами определяется путём усреднения времени работы по всем приборам:

$$T_p = \frac{1}{K} \sum_{i=1}^{K} T_i ,$$

где T_i - время работы прибора $i = \overline{1, K}$;

подставляя последнее выражение в (3.7) окончательно получим:

$$\rho = \lim_{T\to\infty} \frac{1}{KT} \sum_{i=1}^{K} T_i \,;$$

очевидно, что $0 \le \rho \le 1$;

- ***коэффициент простоя*** системы:

$$\eta = 1 - \rho; \qquad (3.8)$$

- ***вероятность потери*** заявок:

$$\pi_n = \lim_{T\to\infty} \frac{N_n(T)}{N(T)}, \qquad (3.9)$$

где T – время работы системы (наблюдения за системой); $N(T)$ – число заявок, поступивших в систему за время T; $N_n(T)$ – число потерянных заявок за время T;

- **вероятность обслуживания** заявки, то есть вероятность того, что поступившая в систему заявка будет обслужена:

$$\pi_0 = (1 - \pi_n) = \lim_{T\to\infty} \frac{N_0(T)}{N(T)}, \qquad (3.10)$$

где $N_0(T)$ – число обслуженных в системе заявок за время T, причем $N_n(T) + N_0(T) = N(T)$ и $\pi_0 + \pi_n = 1$;

- ***производительность*** системы, представляющая собой *интенсивность потока обслуженных заявок*, выходящих из системы:

$$\lambda' = \pi_0 \lambda = (1 - \pi_n)\lambda\,; \qquad (3.11)$$

для *СМО с накопителем неограниченной ёмкости*, при условии отсутствия перегрузок, вероятность потери заявок $\pi_n = 0$ и, следовательно, производительность системы совпадает с интенсивностью поступления заявок в систему: $\lambda' = \lambda$;

- **интенсивность потока потерянных** (не обслуженных) **заявок** из-за ограниченной ёмкости накопителя:

$$\lambda'' = \pi_n \lambda = (1 - \pi_0)\lambda\,; \qquad (3.12)$$

очевидно, что сумма интенсивностей потоков обслуженных и потерянных заявок должна быть равна интенсивности входящего в систему потока заявок: $\lambda' + \lambda'' = \lambda$;

- **среднее время ожидания** заявок в очереди: w;
- **среднее время пребывания** заявок в системе, складывающееся из времени ожидания w и времени обслуживания b:

$$\boxed{u = w + b}; \qquad (3.13)$$

- **средняя длина очереди** заявок:

$$\boxed{l = \lambda' w}; \qquad (3.14)$$

- **среднее число заявок в системе** (в очереди и на обслуживании в *приборе*):

$$\boxed{m = \lambda' u}. \qquad (3.15)$$

Характеристики СМО с неоднородным потоком заявок

Для СМО с неоднородным потоком заявок, в которую поступают H классов заявок с интенсивностями $\lambda_1,\dots,\lambda_H$ и средними длительностями обслуживания $b_1,\dots,b_H$, определяются две группы характеристик обслуживания заявок:

- характеристики по каждому классу (потоку) заявок;
- характеристики объединённого (суммарного) потока заявок.

Характеристики по каждому классу заявок $i=\overline{1,H}$ идентичны характеристикам СМО с однородным потоком:

- *нагрузка*, создаваемая заявками класса i: $y_i = \lambda_i / \mu_i = \lambda_i b_i$;
- *вероятность потери* заявок: π_{n_i};
- *вероятность обслуживания* заявки: $\pi_{0_i} = (1-\pi_{n_i})$;
- *интенсивность* потока *обслуженных* заявок (производительность по i-му классу заявок): $\lambda_{0_i} = \pi_{0_i}\lambda_i = (1-\pi_{n_i})\lambda_i$;
- *интенсивность* потока *потерянных* заявок: $\lambda_{n_i} = \pi_{n_i}\lambda_i$.
- *загрузка* системы, создаваемая заявками класса i:

$\rho_i = \min\left(\dfrac{(1-\pi_{n_i})y_i}{K};1\right)$, где π_{n_i} – вероятность потери заявок класса i из-за ограниченной ёмкости накопителя ($\pi_{n_i}=0$, если ёмкость накопителя – неограниченная); K – число обслуживающих приборов в СМО;

- *время ожидания* заявок в очереди: w_i;
- *время пребывания* заявок в системе: $u_i = w_i + b_i$;
- *длина очереди* заявок: $l_i = \lambda_i w_i$;
- *число заявок в системе* (в очереди и на обслуживании): $m_i = \lambda_i u_i$.

Характеристики объединённого (суммарного) потока заявок позволяют определить усредненные по всем классам заявок показатели эффективности функционирования СМО:

- *суммарная интенсивность* поступления заявок в систему (интенсивность суммарного потока):

$$\Lambda = \sum_{i=1}^{H} \lambda_i ; \qquad (3.19)$$

- *суммарная нагрузка* Y и *суммарная загрузка* R системы:

$$Y = \sum_{i=1}^{H} y_i ; \qquad R = \min(\sum_{i=1}^{H} \rho_i ; 1), \qquad (3.20)$$

причем условие отсутствия перегрузок в СМО с неоднородным потоком заявок и накопителем неограниченной ёмкости имеет вид:

$$R < 1; \qquad (3.21)$$

- *коэффициент простоя* системы: $\eta = 1 - R$;
- *среднее время ожидания* W и *среднее время пребывания* U заявок объединённого потока в системе:

$$W = \sum_{i=1}^{H} \xi_i w_i ; \qquad U = \sum_{i=1}^{H} \xi_i u_i , \qquad (3.22)$$

где $\xi_i = \lambda_i / \Lambda$ – коэффициент, учитывающий долю заявок класса i в суммарном потоке, который может трактоваться как *вероятность того, что поступившая в систему заявка принадлежит классу* i;

- *суммарная длина очереди* и *суммарное число заявок в системе*:

$$L = \sum_{i=1}^{H} l_i ; \qquad M = \sum_{i=1}^{H} m_i . \qquad (3.23)$$

Можно доказать, что для характеристик объединённого (суммарного) потока справедливы те же фундаментальные соотношения (3.13) – (3.15) , что и для однородного потока:

$$U = W + B ; \qquad L = \Lambda W ; \qquad M = \Lambda U ,$$

где B – среднее время обслуживания любой заявки суммарного потока:

$$B = \sum_{i=1}^{H} \xi_i b_i .$$

Характеристики СеМО

Характеристики СеМО делятся на два класса:

- **узловые**, описывающие эффективность функционирования отдельных узлов СеМО;
- **сетевые**, описывающие функционирование СеМО в целом.

Состав *узловых характеристик* СеМО, работающей *в стационарном режиме*, такой же, как и для СМО, и для узла $j = \overline{1, n}$ включает в себя следующие характеристики:

- *нагрузка* узла: $y_j = \lambda_j b_j = \alpha_j \lambda_0 b_j$;
- *загрузка* узла: $\rho_j = \dfrac{y_j}{K_j} = \dfrac{\alpha_j \lambda_0 b_j}{K_j}$, причем $\rho_j < 1$;
- *коэффициент простоя* узла: $\eta_j = 1 - \rho_j$;
- *время ожидания* заявок в узле: w_j;
- *время пребывания* заявок в узле: $u_j = w_j + b_j$;
- *длина очереди* заявок узле: $l_j = \lambda_j w_j = \alpha_j \lambda_0 w_j$;
- *число заявок в* узле (в очереди и на обслуживании):

$m_j = \lambda_j u_j = \alpha_j \lambda_0 (w_j + b_j) = l_j + y_j$.

На основе узловых характеристик рассчитываются *сетевые характеристики* СеМО:

- **суммарная нагрузка** во всех узлах, характеризующая *среднее число заявок, одновременно находящихся на обслуживании во всех узлах сети*:

$$Y = \sum_{j=1}^{n} y_j ,$$

где y_j – нагрузка узла j, причем $0 < Y \le \sum_{j=1}^{n} K_j$;

- **суммарная загрузка** всех узлов СеМО, характеризующая *среднее число параллельно работающих узлов* сети:

$$R = \sum_{j=1}^{n} \rho_j ,$$

где ρ_j – загрузка узла j, причем $0 < R \le n$;

- среднее число заявок, находящихся в очередях всех узлов сети и ожидающих обслуживания:

$$\boxed{L = \sum_{j=1}^{n} l_j}, \qquad (3.26)$$

где l_j – средняя длина очереди заявок в узле j;

- среднее число заявок, находящихся в сети:

$$\boxed{M = \sum_{j=1}^{n} m_j}, \tag{3.27}$$

где m_j – среднее число заявок в узле j, причём для замкнутых сетей это выражение может быть использовано для проверки правильности проведенных расчетов, так как для них число заявок M в сети задано;

- среднее время ожидания заявок в сети:

$$\boxed{W = \sum_{j=1}^{n} \alpha_j w_j}, \tag{3.28}$$

где w_j – среднее время ожидания заявок в узле j; α_j – коэффициент передачи для узла j, показывающий среднее число попаданий заявки в узел j за время её нахождения в сети; $W_j = \alpha_j w_j$ – представляет собой суммарное (полное) время ожидание заявки в узле j за время её нахождения в сети;

- среднее время пребывания заявок в сети:

$$\boxed{U = \sum_{j=1}^{n} \alpha_j u_j}, \tag{3.29}$$

где u_j – среднее время пребывания заявок в узле j; $U_j = \alpha_j u_j$ – суммарное (полное) время пребывания заявки в узле j за время её нахождения в сети;

- **производительность замкнутой СеМО** λ_0, определяемая как интенсивность потока заявок, проходящих через выделенный нулевой узел замкнутой сети, и представляющая собой среднее число заявок, обслуженных в ЗСеМО за единицу времени; производительность ЗСеМО может быть рассчитана на основе выражения (3.5), из которого следует:

$$\boxed{\lambda_0 = \lambda_j / \alpha_j \quad (j = 1, \ldots, n)}; \tag{3.30}$$

Следует отметить, что для сетевых характеристик СеМО выполняются те же фундаментальные соотношения, что и для СМО, а именно:

$$L = \lambda_0 W; \tag{3.31}$$

$$M = \lambda_0 U; \tag{3.32}$$

$$M = L + Y; \tag{3.33}$$

$$U = W + B, \tag{3.34}$$

где $B = \sum_{j=1}^{n} \alpha_j b_j$ – суммарное время обслуживания заявки во всех узлах за время ее нахождения в сети.

Выражения (3.31) и (3.32) представляют собой формулы Литтла для расчёта сетевых характеристик СеМО.

Из (3.32) может быть получена ещё одна важная *формула для расчёта производительности ЗСеМО*:

$$\boxed{\lambda_0 = \frac{M}{U}}. \tag{3.35}$$

Для неоднородной СеМО перечисленные характеристики определяются как для каждого класса в отдельности, так и для объединенного (суммарного) потока заявок.