Моделирование систем массового обслуживания

Системы массового обслуживания

Теория массового обслуживания (или **теория очередей**) имеет дело с процессами, для которых характерна следующая структура.

В **систему массового обслуживания** (СМО) (это могут быть линии связи, приемные пункты, подъездные пути, технологические агрегаты, ремонтные бригады и т. д.) в случайные моменты времени поступают заявки (или требования). Заявки на обслуживание образуют **входной поток**.

Если есть свободные каналы обслуживания, то требование выполняется. Если все каналы обслуживания заняты, то требование становится в очередь по определенным правилам или без обслуживания покидает систему. Выполненные требования образуют **выходной поток**.

Будем считать, что поток требований является простейшим с интенсивностью λ (среднее число требований, поступающих в единицу времени).

СМО состоит из определенного числа обслуживающих единиц – **каналов обслуживания**. Различают одноканальные СМО и многоканальные СМО.

Дисциплина очереди задает порядок прохождения заявки через очередь. Заявки из очереди могут выполняться в порядке поступления, с приоритетом, в случайном порядке и т. д. Очередь может быть конечной или бесконечной. СМО с очередями называют также СМО с ожиданием. Очереди могут ограничиваться по длине (по числу находящихся в ней заявок) или по времени ожидания обслуживания. В СМО с отказом очередь не предусмотрена, то есть заявка, пришедшая в момент, когда заняты все обслуживающие каналы, получает отказ.

Время обслуживания требований в системе является случайной величиной и обычно описывается экспоненциальным (показательным) законом распределения (то есть распределение длительности оставшейся части работ по обслуживанию не зависит от того, сколько оно уже продолжалось) с интенсивностью μ (среднее число требований, выполняемых в единицу времени). Это обусловлено рядом причин:

1) отсутствием последействия;
2) простотой и удобством аналитических выражений;
3) именно так устроены многие реальные системы.

Показательное распределение времени обслуживания имеет вид: $P_t = \mu e^{-\mu t}$ $(t \geq 0)$. Тогда среднее время обслуживания одним каналом одного требования $t_{обсл} = \dfrac{1}{\mu}$.

Коэффициент загрузки СМО (среднее число каналов, которое должно быть для обслуживания в единицу времени всех поступающих требований) $\rho = \dfrac{\lambda}{\mu}$.

Одноканальная СМО с отказами

СМО содержит один обслуживающий канал. На вход поступает простейший поток заявок с интенсивностью $\boldsymbol{\lambda}$. Образование очереди не допускается. Если заявка застала обслуживающий канал занятым, то она покидает систему.

Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $t_{обсл} = 1/\mu$.

Возможные состояния СМО S_0 (канал свободен) и S_1 (канал занят).

Размеченный граф состояний одноканальной СМО с отказами имеет следующий види (рис.7.1):

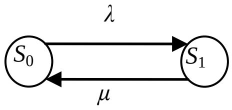

Рис. 7.1. Граф состояний одноканальной СМО с отказами

Показатели эффективности работы СМО:

1) вероятность отказа $p_{отк}$ (вероятность того, что заявка покинет СМО необслуженной, т.е. предельная вероятность состояния S_1)

$$p_{отк} = p_1;$$

2) относительная пропускная способность Q (отношение среднего числа обслуживаемых в единицу времени заявок к среднему числу поступивших за это время заявок)

$$Q = 1 - p_{отк};$$

3) абсолютная пропускная способность A (среднее число заявок, которое СМО может обслужить в единицу времени)

$$A = \lambda Q.$$

Пример. Одноканальная телефонная линия. Заявка-вызов, поступившая в момент, когда линия занята, получает отказ. Простейший поток заявок поступает с интенсивностью λ=50 звонков/ч. Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения. Средняя продолжительность разговора $t_{обсл}$ = 3 мин. Определим показатели эффективности работы СМО.

Решение.

Данная телефонная линия – это одноканальная СМО с отказами. Время обслуживания $t_{обсл}$ = 3 мин = 3/60 ч = 0,05 ч. Тогда интенсивность обслуживания μ = $1/t_{обсл}$ = 1/0,05 = 20 звонков/ч. Размеченный граф состояний имеет следующий вид (рис. 7.2):

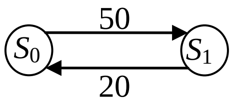

Рис. 7.2. Граф состояний

Пусть p_0 – предельная вероятность состояния S_0. Состояние S_1 связано с состоянием S_0 двумя стрелками с интенсивностями 50 и 20. Пусть p_1 – предельная вероятность состояния S_1. Тогда

$$p_1 = \frac{50}{20} p_0 = 2{,}5 p_0.$$

Так как $p_0 + p_1 = 1$, то $1 = p_0 + 2{,}5p_0 = 3{,}5p_0$. Отсюда $p_0 = 0{,}286$. Тогда $p_1 = 2{,}5p_0 = 0{,}714$.

Вероятность отказа $p_{отк}$ – это вероятность того, что линия занята, то есть предельная вероятность состояния S_1. Поэтому $p_{отк} = p_1 = 0{,}714$.

Относительная пропускная способность $Q = 1 - p_{отк} = 1 - 0{,}714 = 0{,}286$. Это вероятность того, что заявка будет обслужена.

Абсолютная пропускная способность $A = \lambda Q = 50 \cdot 0{,}286 = 14{,}3$ звонка/ч, то есть в среднем в час СМО обслуживает 14,3 звонка.

Мы видим, что номинальная пропускная способность телефонной линии μ=20 звонков/ч отличается от абсолютной пропускной способности A = 14,3 звонка/ч из-за случайного характера потока звонков и случайности времени обслуживания.

Многоканальная СМО с отказами

СМО содержит n обслуживающих каналов. На вход поступает простейший поток заявок с интенсивностью λ. Образование очереди не допускается. Если заявка застала все обслуживающие каналы занятыми, то она покидает систему. Если в момент поступления требования имеется свободный канал, то он немедленно приступает к обслуживанию поступившего требования. Каждый канал может одновременно обслуживать только одно требование. Все каналы функционируют независимо. Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $t_{обсл} = 1/\mu$.

Возможные состояния СМО S_0 (все каналы свободны), S_1 (один канал занят, остальные свободны), S_2 (два канала заняты, остальные свободны), , S_n (все каналы заняты).

Размеченный граф состояний многоканальной СМО с отказами имеет следующий вид (рис.7.3):

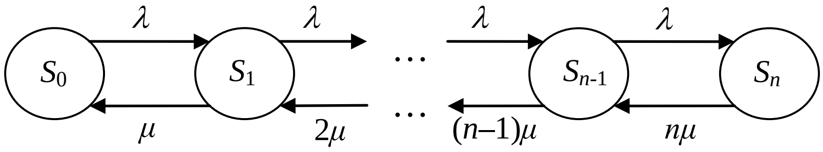

Рис. 2.3. Граф состояний системы

Приведенная интенсивность потока заявок (интенсивность нагрузки канала) $\rho = \lambda/\mu$.

Показатели эффективности работы СМО:

1) p_0 (вероятность того, что все обслуживающие каналы свободны);

2) вероятность отказа $p_{отк}$ (вероятность того, что заявка покинет СМО необслуженной) $p_{отк}=p_n$;

3) p_k (вероятность того, что в системе k требований) $p_k = \frac{\rho^k}{k!} p_0$;

4) относительная пропускная способность Q (отношение среднего числа обслуживаемых в единицу времени заявок к среднему числу поступивших за это время заявок) $Q=1–p_{отк}$;

5) абсолютная пропускная способность A (среднее число заявок, которое СМО может обслужить в единицу времени) $A=\lambda Q$;

6) среднее число свободных от обслуживания каналов N_0 есть математическое ожидание числа свободных каналов $N_0=np_0+(n–1)p_1+\ldots+1p_{n-1}+0p_n$;

7) коэффициент простоя каналов $K_{пр} = \frac{N_0}{n}$;

8) среднее число занятых обслуживанием каналов $N_{зан}=\rho Q$; $n=N_0+N_{зан}$;

9) коэффициент загрузки каналов $K_{зан} = \frac{N_{зан}}{n}$.

Пример. Трехканальная телефонная линия. Заявка-вызов, поступившая в момент, когда все каналы заняты, получает отказ. Простейший поток заявок с интенсивностью λ=60 звонков/ч. Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения. Средняя продолжительность разговора 3 мин. Определить показатели эффективности работы СМО.

Решение.

Данная телефонная линия – это многоканальная СМО с отказами.

$t_{обсл}$=3 мин=0,05 ч. Тогда интенсивность обслуживания $\mu=1/t_{обсл}=1/0{,}05=20$ звонков/ч.

Коэффицент загрузки СМО $\rho = \frac{\lambda}{\mu} = \frac{60}{20} = 3$.

Размеченный граф состояний имеет следующий вид (рис. 7.4):

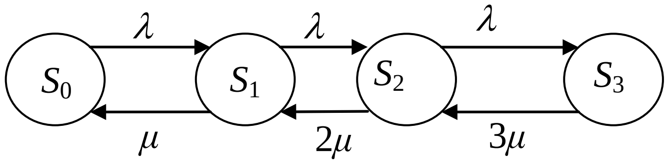

Рис. 7.4. Граф состояний системы

p_0 – предельная вероятность состояния S_0.

p_1 – предельная вероятность состояния S_1. Имеем $p_1 = \frac{\lambda}{\mu} p_0 = \rho p_0 = 3p_0$.

Аналогично p_2 – предельная вероятность состояния S_2 $p_2 = \frac{\lambda}{2\mu} p_1 = \frac{\rho}{2} p_1 = \frac{3}{2} \cdot 3p_0 = 4,5p_0$

,

p_3 – предельная вероятность состояния S_3 $p_3 = \frac{\lambda}{3\mu} p_2 = \frac{\rho}{3} p_2 = \frac{3}{3} \cdot 4,5p_0 = 4,5p_0$.

Так как $p_0+p_1+p_2+p_3=1$, то $1= p_0+3p_0+4,5p_0+4,5p_0=13p_0$. Отсюда $p_0=0,077$ (вероятность того, что все обслуживающие каналы свободны). Тогда $p_1=3\cdot 0,077=0,231$, $p_2=4,5\cdot 0,077=0,346$, $p_3=4,5\cdot 0,077=0,346$.

Вероятность отказа $p_{отк}$ – это вероятность того, что все каналы заняты, то есть предельная вероятность состояния S_3. Поэтому $p_{отк}=p_3=4,5\cdot 0,077=0,346$.

Относительная пропускная способность $Q=1–p_{отк}=1–0,346=0,654$. Это вероятность того, что заявка будет обслужена.

Абсолютная пропускная способность $A=\lambda Q=60\cdot 0,654=39,24$ звонков/ч, то есть в среднем в час СМО обслуживает 39,24 звонка.

Среднее число свободных от обслуживания каналов N_0 есть математическое ожидание числа свободных каналов:

$$N_0=3p_0+2p_1+1p_2+0p_3=3\cdot 0,077+2\cdot 0,231+0,346=1,039.$$

Коэффицент простоя каналов $K_{пр} = \frac{N_0}{n} = \frac{1,039}{3} = 0,346$.

Среднее число занятых обслуживанием канлов $N_{зан}=\rho Q=3\cdot 0,645=1,962$.

Коэффицент загрузки каналов $K_{зан} = \frac{N_{зан}}{n} = \frac{1,962}{3} = 0,654$.

Одноканальная СМО с неограниченной очередью

В этом случае клиенты формируют одну очередь к единственному пункту обслуживания. Пусть

λ – число заявок в единицу времени;

μ – число клиентов, обслуживаемых в единицу времени;

n – число заявок в системе.

Возможные состояния СМО S_0 (канал свободен), S_1 (канал занят, очереди нет), S_2 (канал занят, в очереди одна заявка), S_3 (канал занят, в очереди две заявки) и т.д.

Размеченный граф состояний одноканальной СМО с неограниченной очередью имеет следующий вид (рис. 7.5):

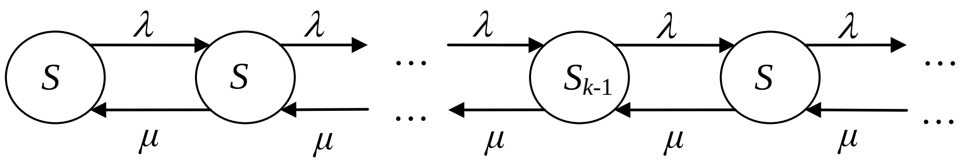

Рис. 7.5. Граф состояний одноканальной СМО с неограниченной очередью

Формулы для описания системы:

$L_{сист} = \frac{\lambda}{\mu - \lambda} = \frac{\rho}{1-\rho}$ – среднее число клиентов в системе;

$T_{сист} = \frac{1}{\mu - \lambda} = \frac{L_{сист}}{\lambda}$ – среднее время обслуживания одного клиента в системе (время ожидания в очереди плюс время обслуживания);

$L_{обсл} = \rho$ – среднее число заявок, находящихся под обслуживанием;

$L_{оч} = L_{сист} - L_{обсл} = \dfrac{\lambda^2}{\mu(\mu-\lambda)} = \dfrac{\rho^2}{1-\rho}$ – среднее число клиентов в очереди – средняя длина очереди;

$T_{оч} = \dfrac{\lambda}{\mu(\mu-\lambda)} = \dfrac{L_{оч}}{\lambda}$ – среднее время ожидания клиента в очереди;

$p_0 = 1 - \dfrac{\lambda}{\mu}$ – вероятность отсутствия заявок в системе;

$p_{зан} = 1 - p_0 = \rho$ – вероятность того, что канал занят;

$p_k = \rho^k(1-\rho)$ – вероятность того, что в системе ровно k клиентов;

$p_{n>k} = \rho^{k+1}$ – вероятность того, что в системе находится более чем k клиентов.

Пример. Магазин с одним продавцом. Предполагается, что простейший поток покупателей поступает с интенсивностью 20 чел/ч. Время обслуживания заявки – случайная величина, которая подчиняется экспоненциальному закону распределения с параметром равным 25 чел/ч. Определить:

1) среднее время пребывания покупателя в очереди;
2) среднюю длину очереди;
3) среднее число покупателей в магазине;
4) среднее время пребывания покупателя в магазине;
5) вероятность того, что в магазине не окажется покупателей;
6) вероятность того, что в магазине окажется ровно 4 покупателя.

Решение.

Данный магазин – одноканальная СМО с неограниченной очередью и коэффициентом загрузки $\rho = \dfrac{\lambda}{\mu} = \dfrac{20}{25} = 0,8$.

Вероятность того, что в магазине не окажется покупателей, равна $p_0 = 1 - \dfrac{\lambda}{\mu} = 1 - 0,8 = 0,2$.

Вероятность того, что в магазине окажется ровно 4 покупателя, равна

$$p_4 = \rho^4(1-\rho) = 0,8^4 \cdot (1-0,8) = 0,082.$$

Средняя длина очереди $L_{оч} = \dfrac{\rho^2}{1-\rho} = \dfrac{0,8^2}{1-0,8} = 3,2$.

Среднее время пребывания покупателя в очереди $T_{оч} = \dfrac{L_{оч}}{\lambda} = \dfrac{3,2}{20} = 0,16$ ч=9,6 мин.

Среднее число покупателей в магазине $L_{сист} = \dfrac{\rho}{1-\rho} = \dfrac{0,8}{1-0,8} = 4$.

Среднее время пребывания покупателя в магазине $T_{сист} = \dfrac{L_{сист}}{\lambda} = \dfrac{4}{20} = 0,2$ ч=12 мин.

Многоканальная СМО с неограниченной очередью

В многоканальной системе для обслуживания открыты два канала или более. Предполагается, что клиенты ожидают в общей очереди и обращаются в первый освободившийся канал обслуживания. Пример такой многоканальной однофазовой системы можно увидеть во многих банках: из общей очереди клиенты обращаются в первое освободившееся окошко для обслуживания.

В многоканальной системе поток заявок подчиняется *пуассоновскому* закону с параметром λ, а время обслуживания – *экспоненциальному* с параметром μ. Приходящий первым обслуживается первым, и все каналы обслуживания работают в одинаковом темпе.

Возможные состояния СМО S_0 (все каналы свободны), S_1 (один канал занят, остальные свободны), S_2 (два канала заняты, остальные свободны), , S_n (все каналы заняты), S_{n+1} (все каналы заняты, в очереди одна заявка), S_{n+2} (все каналы заняты, в очереди две заявки) и т.д.

Размеченный граф состояний многоканальной СМО с неограниченной очередью имеет следующий вид (рис. 7.6):

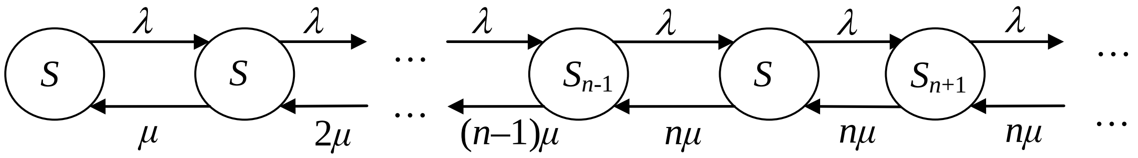

Рис. 7.6. Граф состояний системы

Формулы для описания системы:

$$p_0=\left(1+\frac{\rho}{1!}+\frac{\rho^2}{2!}+\ldots+\frac{\rho^n}{n!}+\frac{\rho^{n+1}}{n!(n-\rho)}\right)^{-1}$$ – вероятность того, что система свободна;

$$p_n=\frac{\rho^n}{n!}p_0$$ – вероятность того, что в системе находится n заявок;

$$p_{n+1}=\frac{\rho}{n}\cdot\frac{\rho^n}{n!}p_0;\quad p_{n+2}=\left(\frac{\rho}{n}\right)^2\cdot\frac{\rho^n}{n!}p_0;\quad p_{n+3}=\left(\frac{\rho}{n}\right)^3\cdot\frac{\rho^n}{n!}p_0$$ и т.д.

$$p_q=\frac{\rho^{n+1}}{n!(n-\rho)}p_0$$ – вероятность того, что заявка окажется в очереди;

$$L_{оч}=\frac{\rho^{n+1}p_0}{n!n(1-\frac{\rho}{n})^2}$$ – среднее число заявок в очереди;

$$L_{сис}=L_{оч}+\rho$$ – среднее число заявок в системе;

$$T_{оч}=\frac{1}{\lambda}L_{оч}$$ – среднее время нахождения заявки в очереди;

$$T_{сис}=\frac{1}{\lambda}L_{сис}$$ – среднее время нахождения заявки в системе.

Пример. Магазин с двумя продавцами. Предполагается, что простейший поток покупателей поступает с интенсивностью 20 чел/ч. Время обслуживания заявки – случайная величина, которая подчиняется экспоненциальному закону распределения с параметром равным 25 чел/ч. Определить показатели эффективности СМО.

Решение.

Данный магазин – двухканальная СМО с неограниченной очередью и коэффициентом загрузки $\rho=\frac{\lambda}{\mu}=\frac{20}{25}=0{,}8$.

Вероятность того, что в магазине не окажется покупателей, равна

$$p_0=\left(1+\frac{\rho}{1!}+\frac{\rho^2}{2!}+\frac{\rho^3}{2!(2-\rho)}\right)^{-1}=\left(1+\frac{0{,}8}{1!}+\frac{0{,}8^2}{2!}+\frac{0{,}8^3}{2!(2-0{,}8)}\right)^{-1}=0{,}429.$$

Вероятность того, что в магазине окажется ровно 4 покупателя (то есть 2 покупателя обслуживаются и еще 2 покупателя в очереди), равна

$$p_{n+2}=\left(\frac{\rho}{n}\right)^2\cdot\frac{\rho^n}{n!}p_0=\left(\frac{0{,}8}{2}\right)^2\cdot\frac{0{,}8^2}{2!}\cdot 0{,}429=0{,}022.$$

Среднее число заявок в очереди $L_{оч} = \dfrac{\rho^{n+1} p_0}{n!n(1-\frac{\rho}{n})^2} = \dfrac{0,8^3 \cdot 0,429}{2!\cdot 2(1-\frac{0,8}{2})^2} = 0,153$.

Среднее время пребывания покупателя в очереди $T_{оч} = \dfrac{1}{\lambda} L_{оч} = \dfrac{0,153}{20} = 0,008$ ч=0,48 мин.

Среднее число покупателей в магазине $L_{сис} = L_{оч} + \rho = 0,153 + 0,8 = 0,953$.

Среднее время пребывания покупателя в магазине $T_{сис} = \dfrac{1}{\lambda} L_{сис} = \dfrac{0,953}{20} = 0,048$ ч=2,88 мин.

Одноканальная СМО с ограниченной очередью

СМО содержит один обслуживающий канал. На вход поступает простейший поток заявок с интенсивностью λ. Если заявка застала обслуживающий канал занятым, то она встает в очередь и ожидает начала обслуживания. Число мест в очереди ограничено и равно m. Если заявка застала обслуживающий канал занятым и в очереди нет свободных мест, то она покидает систему необслуженной.

Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $t_{обсл}=1/\mu$.

Возможные состояния СМО S_0 (канал свободен), S_1 (канал занят, очереди нет), S_{1+1} (канал занят, в очереди одна заявка), S_{1+2} (канал занят, в очереди две заявки),…, S_{1+m} (канал занят, в очереди m заявок).

Размеченный граф состояний многоканальной СМО с неограниченной очередью имеет следующий вид (рис. 7.7):

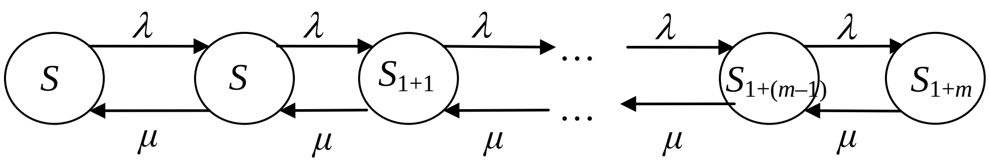

Рис. 7.7. Граф состояний системы

Формулы для описания системы:

1) $p_0 = \dfrac{1-\rho}{1-\rho^{m+2}}$ – вероятность того, что канал свободен;

2) $p_1=\rho p_0$;

3) $p_{1+k}=\rho^{1+k} p_0$ – вероятность того, что канал занят, в очереди k заявок;

4) $p_{отк}=p_{1+m}= \rho^{1+m} p_0$ – вероятность отказа (канал занят, в очереди нет свободных мест);

5) относительная пропускная способность Q (отношение среднего числа обслуживаемых в единицу времени заявок к среднему числу поступивших за это время заявок) $Q=1–p_{отк}$;

6) абсолютная пропускная способность A (среднее число заявок, которое СМО может обслужить в единицу времени) $A=\lambda Q$;

7) среднее число заявок в очереди $L_{оч} = \rho^2 \dfrac{1-\rho^m(m+1-m\rho)}{(1-\rho^{m+2})(1-\rho)}$;

8) среднее время нахождения заявки в очереди $T_{оч} = \dfrac{1}{\lambda} L_{оч}$;

9) среднее число заявок, находящихся под обслуживанием (среднее число занятых каналов)

$$L_{обсл}=1-p_0;$$

10) среднее число заявок в системе $L_{сис}=L_{оч}+L_{обсл}$;

11) среднее время нахождения заявки в системе $T_{сис}=\frac{1}{\lambda}L_{сис}$.

Пример. Автозаправочная станция имеет одну бензоколонку с площадкой, допускающей пребывание в очереди на заправку не более трех автомашин одновременно. Если в очереди находятся две автомашины, то очередная прибывшая автомашина проезжает мимо автозаправочной станции. Предполагается, что простейший поток автомашин поступает на станцию с интенсивностью 10 автомашин/ч. Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром 12 автомашин/ч. Определить параметры системы.

Решение.

Данная автозаправочная станция – это одноканальная СМО с ограниченной очередью с коэффициентом загрузки $\rho=\frac{\lambda}{\mu}=\frac{10}{12}=\frac{5}{6}$.

Вероятность того, что на станции нет автомашин, равна $p_0=\frac{1-\rho}{1-\rho^{m+2}}=\frac{1-\frac{5}{6}}{1-\left(\frac{5}{6}\right)^{2+2}}=0,322$.

$p_1=\rho p_0=5/6\cdot 0,322=0,268$; $p_{1+1}=\frac{\lambda}{\mu}p_1=\frac{5}{6}\cdot 0,268=0,224$; $p_{1+2}=\frac{\lambda}{\mu}p_2=\frac{5}{6}\cdot 0,224=0,186=p_{отк}$.

Относительная пропускная способность $Q=1-p_{отк}=1-0,186=0,814$. Это вероятность того, что заявка будет обслужена.

Абсолютная пропускная способность $A=\lambda Q=10\cdot 0,814=8,14$ автомашин/ч.

Среднее число автомашин в очереди

$$L_{оч}=\rho^2\frac{1-\rho^m(m+1-m\rho)}{(1-\rho^{m+2})(1-\rho)}=\left(\frac{5}{6}\right)^2\frac{1-\left(\frac{5}{6}\right)^2\left(2+1-2\cdot\frac{5}{6}\right)}{\left(1-\left(\frac{5}{6}\right)^{2+2}\right)\left(1-\frac{5}{6}\right)}=0,596.$$

Среднее время нахождения автомашины в очереди $T_{оч}=\frac{1}{\lambda}L_{оч}=0,596/10=0,0596$ ч=3,576 мин.

Среднее число заявок, находящихся под обслуживанием (среднее число занятых каналов)

$$L_{обсл}=1-p_0=1-0,322=0,678.$$

Среднее число автомашин на станции $L_{сис}=L_{оч}+L_{обсл}=0,596+0,678=1,274$ автомашины.

Среднее время нахождения автомашины на станции

$$T_{сис}=\frac{1}{\lambda}L_{сис}=1,274/10=0,1274 \text{ ч}=7,644 \text{ мин}.$$