

ЗАДАЧА О ВРЕМЕНИ ОЖИДАНИЯ

Е. Н. Бухман

(Москва)

Предприятия транспорта, электростанций, связи и др., осуществляющие обслуживание массового потребителя, сталкиваются с необходимостью расчета технических средств в условиях неравномерной нагрузки, подверженной постоянным колебаниям как периодическим, так и нерегулярным.

Одной из основных и наиболее распространенных характеристик качества обслуживания является время ожидания. Отсюда возникает задача определения среднего времени ожидания γ или вероятности ожидания выше заданного срока $P(>\tau)$, в зависимости от величины потока нагрузки u и от числа обслуживателей s .

Решению этой задачи посвящено значительное число исследований, рассматривающих обычно отдельные частные случаи. Укажем некоторые из важнейшие из них.

В 1917 г. датским математиком Эрлангом^[1] были определены $P(>\tau)$ и γ при экспоненциальном распределении времени занятия t , а также при постоянстве времени занятия t , но для значений s не больше 3.

Позднее американский математик Молина^[2] подтвердил решение Эрланга при экспоненциальном распределении t и дал приближенное решение при постоянном t .

Изящное решение той же задачи при экспоненциальном распределении t дано А. Н. Колмогоровым^[3]. В 1930 г. Экелеф^[4] получил приближенное решение также для случая постоянного времени занятия.

Американскому исследователю Кроммелину^[5] удалось получить точное решение задачи при постоянном времени занятия, но в весьма сложной аналитической форме, трудной для осуществления вычислений. В 1932 г. А. Я. Хинчин^[6] дал формулу среднего времени ожидания γ и метод расчета моментов распределения $P(>\tau)$ при любом законе распределения t , но для случая одного обслуживателя $s = 1$.

Наиболее широкие обобщения дал Поллячен^[7], исследовавший задачу при любом числе линий и любом законе распределения t . Полученные им в весьма сложной аналитической форме решения оказалось возможным привести к виду, пригодному для практического использования путем числового расчета лишь в некоторых частных случаях. При экспоненциальном распределении t решение Поллячека подтверждает решение Эрланга и Молина, при постоянстве t его решение совпадает с решением Кроммелина, при $s = 1$ его вывод совпадает с выводом Хинчина. Однако для $s > 1$ при любом распределении t Полляченку не удалось найти пригодного практически решения.

В настоящей статье излагается решение задачи для любого распределения времени занятия t , точное при достаточно большом s и имеющее простую аналитическую форму. Это решение дано применительно к работе автоматической телефонной станции.

1. Рассматриваем процесс обслуживания неограниченно большого числа абонентов группой s соединительных устройств (линий), каждое из которых в одинаковой мере доступно каждому абоненту. Обслуживание вызовов предполагается происходящим строго в порядке поступления вызовов. Вызовы считаем случайными и в коллективном и в индивидуальном случаях (см. Фрай^[8]) и потому вероятность поступления x вызовов за время t определяющейся по формуле Пуассона.

Группу линий принимаем находящейся в «статистическом равновесии». Это означает, что вероятность определенного состояния группы в отношении числа занятых линий или числа абонентов, ожидающих обслуживания, не зависит от времени. В то же время это означает, что вероятность перехода группы от состояния n занятых линий к $n+1$ занятых линий (или от m ожидающих к $m+1$ ожидающих) за некоторый промежуток времени Δt равен вероятности обратного перехода от $n+1$ занятых линий к n (соответственно от $m+1$ ожидающих к m) за тот же промежуток времени Δt .

Исходя из этого условия, наиболее простым путем можно определить величину вероятности P_n одновременной занятости n линий или вероятности W_m наличия m ожидающих при занятости всех s линий. В математической форме это условие может быть выражено, как

$$P_n p^+ = P_{n+1} p_{n+1}^-, \quad W_m p^+ = W_{m+1} p_s^- \quad (1.1)$$

Здесь через p^+ обозначена вероятность возникновения одного вызова на протяжении малого промежутка времени Δt . В силу принятых условий она не зависит ни от состояния групп линий, ни от моментов начала занятых отдельных линий от переговоров.

Через p_{n+1}^- обозначена вероятность освобождения одной линии из $n+1$ занятых на протяжении того же малого промежутка времени Δt . Очевидно, что эта вероятность зависит от числа занятых линий, не могущего превысить s , но не от числа ожидающих.

Вероятность p^+ может быть найдена следующим путем. Согласно условию о случайному характере потока вызовов вероятность поступления одного вызова за время Δt при математическом ожидании числа их y за время одного занятия t^* есть

$$p^+ = y \frac{\Delta t}{t^*} e^{-y \Delta t / t^*} = \frac{y}{t^*} \Delta t - \left(\frac{y}{t^*} \right)^2 \Delta t^2 + \frac{1}{2} \left(\frac{y}{t^*} \right)^3 \Delta t^3 - \dots$$

Здесь использовано разложение экспоненциальной функции в ряд. При малом значении Δt имеем

$$p^+ \approx \frac{y}{t^*} \Delta t \quad (1.2)$$

Вероятность p_n^- очевидно, определяется как pr_1^- , будучи пропорциональной числу занятых линий n .

Вероятность освобождения единичной линии p_1^- находим следующим путем. Обозначим через $p(t) dt$ вероятность того, что наудачу выбранный переговор, наблюдаемый в момент его возникновения, имеет длительность от t до $t+dt$. Оценка возможной длительности переговора изменяется, если этот переговор наблюдается не в момент его возникновения, а в момент, когда он уже происходил в течение некоторого времени до момента наблюдения. Наблюдая в произвольный момент времени отдельные наличные в этот момент переговоры, мы будем чаще наблюдать переговоры более высокой, чем малой, длительности. Вероятность $p'(t) dt$ застать налицо переговор длительности от t до $t+dt$ по сравнению с вероятностью $p(t) dt$ тем больше, чем больше t . Таким образом,

$$p'(t) = \frac{t}{t^*} p(t) \quad (1.3)$$

Коэффициент $1/t^*$ возникает из условия, что

$$\int_0^\infty p'(t) dt = 1$$

В данном случае имеет место полная аналогия с определением той или иной длительности жизни для новорожденного и для человека, наудачу выбранного из общей массы живущих на определенный момент времени. Среди массы одновременно живущих естественным путем отбираются индивидуумы наибольшей долговечности, и большей вероятностью попасть в состав этой случайной выборки из состава живущих обладает тот, кто обладает и наибольшим долголетием, так как из числа преждевременно погибающих к моменту производства выборки значительное число уже отселялось. Поэтому соотношение распределений продолжительности жизни для новорожденных и для наличной в некоторый момент массы людей отвечает соотношению между вероятностями $p(t)$ и $p'(t)$.

То же самое имеет место для распределений по срокам службы массового имущества, вновь вступающего в эксплуатацию и наличного, см. [19].

Вероятность того, что переговор длительностью t закончится в течение промежутка Δt , очевидно, определяется как $\Delta t/t$, поскольку вероятность застать переговор в любой стадии его осуществления от 0 до t всегда одна и та же для любого промежутка одной и той же длины на всем протяжении отрезка t . Вероятность того, что наудачу взятый переговор имеет длительность t^* и заканчивается в течение промежутка Δt , есть

$$\frac{\Delta t}{t} p'(t) dt = \frac{\Delta t}{t^*} \frac{t}{t^*} p(t) dt = \frac{\Delta t}{t^*} p(t) dt$$

Полная вероятность того, что переговор любой длительности окончится за промежуток Δt , есть

$$p_1^- = \frac{\Delta t}{t^*} \int_0^\infty p(t) dt = \frac{\Delta t}{t^*}$$

Глажко отметить, что эта вероятность не зависит от распределения времени занятия переговором t , а только от среднего времени занятия t^* и промежутка Δt . В результате имеем

$$p_n^- = n \frac{\Delta t}{t^*} \quad (1.4)$$

Определив элементарные вероятности p^+ и p_n^- , подстановкой их в равенство (1.2) получаем рекуррентные формулы

$$P_{n+1} = P_n \frac{y}{n+1}, \quad W_{m+1} = W_m \frac{y}{s}$$

Отсюда, учитывая, что $W_0 = P_s$, находим

$$P_n = P_0 \frac{y^n}{n!}, \quad W_m = P_0 \frac{y^s}{s!} \left(\frac{y}{s}\right)^m \quad (1.5)$$

Величина P_0 легко может быть определена из условия, что

$$\sum_{n=0}^{s-1} P_n + \sum_{m=0}^{\infty} W_m = 1$$

Эти формулы достаточно хорошо известны для случая экспоненциального распределения t . Приведенный здесь вывод показывает пригодность этих формул при любом законе распределения t в силу того, что элементарные вероятности p^+ и p_n^- оказываются независящими от закона распределения t .

Распределение моментов окончания занятых при достаточно большом s подчиняется закону Пуассона. Это означает, что вероятность окончания m переговоров за время τ при $a\tau$ переговорах, заканчивающихся в среднем за время τ , определяется как

$$P(m) = \frac{(a\tau)^m}{m!} e^{-a\tau} \quad (1.6)$$

В самом деле, если для отдельной линии вероятность окончания ее занятия за время τ обозначим через p^- , то вероятность того, что при наличии s линий за время τ освободятся m линий (т. е. закончится m занятий), определяется по известной формуле Бернулли как

$$P(m) = C_s^m [p^-]^m [1 - p^-]^{s-m}$$

Применение этой формулы означает, что использование абонентом занятой им линии не зависит от использования остальных линий, т. е. что вероятности p^- для отдельных линий независимы между собой. Это предположение представляется вполне законным, особенно если принять во внимание, что в данном случае каждая освобождающаяся линия оказывается немедленно занятой ближайшим по очереди абонентом.

При достаточно большом s формула Бернулли, как известно, преобразуется в формулу Пуассона, написанную выше, если заменить через $a\tau$ произведение, $s p^-$ выражющее математическое ожидание переменной m .

При малых значениях s закон распределения моментов окончания занятых выражается иначе. Тем не менее решение, получающееся при использовании формулы Пуассона, дает хороший результат уже при $s=5$, причем среднее время ожидания γ определяется вполне точно даже при $s=1$.

При выводе формулы Эрланга-Молина математическое ожидание числа занятых a определяется как отношение s / t^* . Это выражение, как увидим ниже, справедливо только при экспоненциальном распределении времени занятия t . Поэтому формула Эрланга отвечает только частному случаю экспоненциального распределения t . Для определения a при любом законе распределения t нужно найти математическое ожидание b времени окончания занятия для любого переговора, наблюдавшегося в произвольный момент времени. Эта величина, очевидно, равна половине среднего времени переговора, но определяемого не на основе распределения $p(t)$, как это нередко делается, а на основе распределения $p'(t)$, так как речь идет о средней длительности переговора в произвольный момент его осуществления.

В самом деле, при наблюдении в произвольный момент времени одинаково вероятно застать ведущийся переговор в любой стадии его осуществления. Поэтому вероятность того, что данному переговору длительностью t остается время τ , есть τ / t . При заданном t математическое ожидание

$$\tau^* = \int_0^t \frac{\tau}{t} d\tau = \frac{t}{2}$$

При любом t математическое ожидание окончания переговора θ есть

$$\theta = \int_0^\infty \frac{t}{2} p'(t) dt = \int_0^\infty \frac{t^2}{t^*} p(t) dt = \frac{1}{2} \frac{m_2}{t^*} = \frac{1}{2} t^* (1 + V^2) \quad (1.7)$$

где m_2 — второй начальный момент и V — коэффициент вариации распределения $p(t)$.

Очевидно, что математическое ожидание числа занятых, оканчивающихся за единицу времени из числа наблюдаемых в произвольный момент времени для одного только соединительного устройства, есть величина, обратная θ , т. е. $2t^*/m_2$ согласно (1.7), а для всех s соединительных устройств

$$a = s \frac{2t^*}{m_2} = \frac{s}{\theta} \quad (1.8)$$

Вероятность $p_m (> \tau)$ того, что за время τ закончится не менее чем m вызовов и, следовательно, длительность ожидания будет не менее, чем τ , определяется на основе найденной ранее вероятности $P(m)$ объединением вероятностей всех возможных значений числа оканчивающихся переговоров от 0 до m (переменную суммирования обозначаем через y):

$$p_m (> \tau) = \sum_{y=0}^m \frac{(a\tau)^y}{y!} e^{-a\tau} \quad (1.9)$$

Полная вероятность ожидания не меньше, чем τ , при любом m есть

$$P(>\tau) = \sum_{m=0}^{\infty} W(m) p_m (> \tau) \quad (1.10)$$

Подставляя значения $W(m)$ по (1.5) и $p_m (> \tau)$ по (1.9), имеем

$$P(>\tau) = \sum_{m=0}^{\infty} P_0 \frac{y^s}{s!} \left(\frac{y}{s} \right)^m \sum_{y=0}^m \frac{(a\tau)^y}{y!} e^{-a\tau}$$

Отсюда путем элементарных преобразований получаем

$$P(>\tau) = P(>0) e^{-(s-y)\tau / \theta} \quad (P(>0) = P_0 \frac{y^s}{s!} \frac{s}{s-y}) \quad (1.11)$$

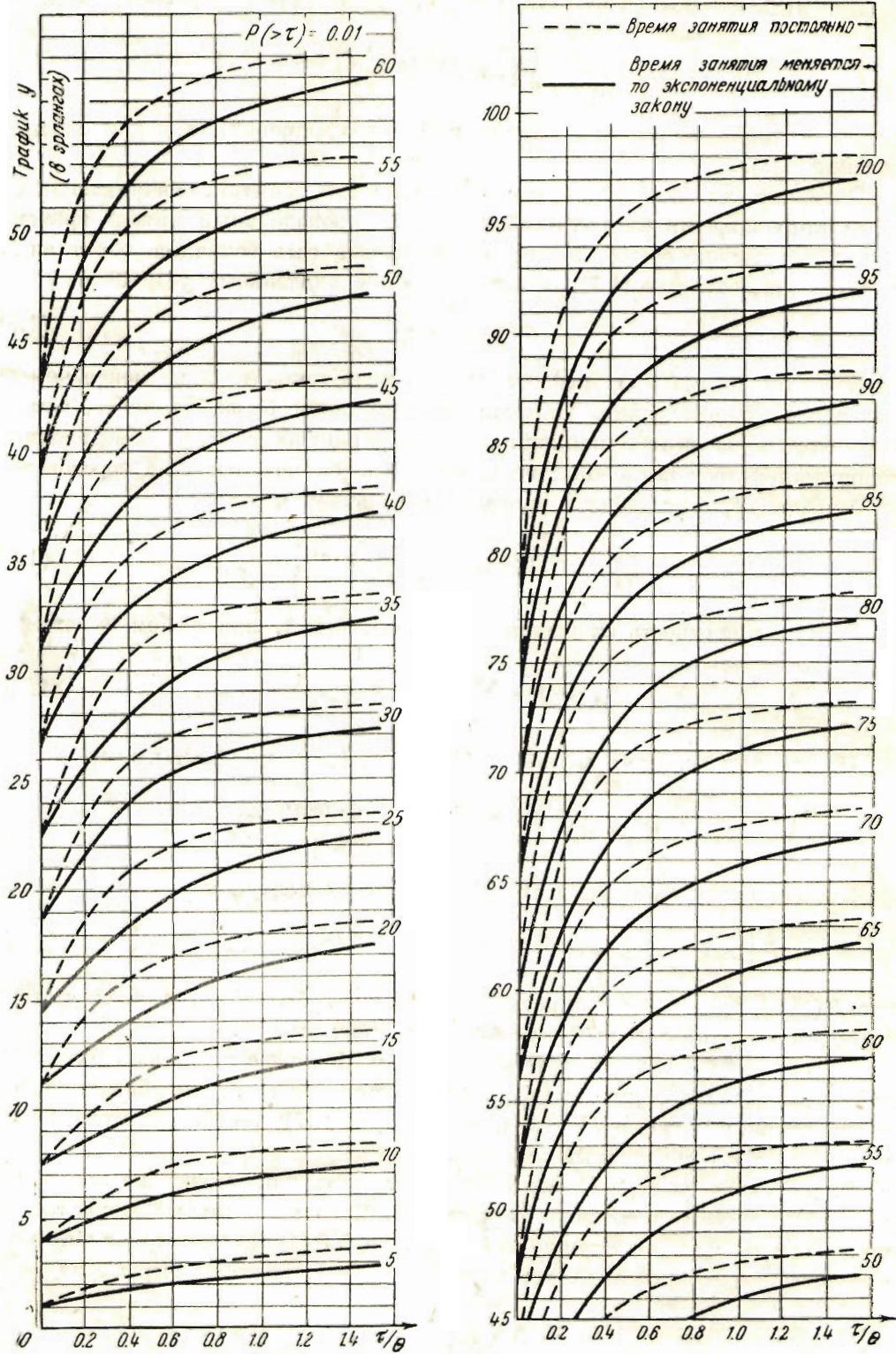
где $P(>0)$ представляет вероятность ожидания больше 0, т. е. отсутствия свободных соединительных устройств в момент вызова.

Такого вида формула дана Эрлангом и Молина с той лишь разницей, что вместо θ в их формулах фигурирует величина t^* ; при экспоненциальном распределении t , когда $V=1$ и на основе (1.7) $\theta=t^*$, формула (1.11) приобретает именно данный ими вид.

2. Вероятность $P(>\tau)$ можно рассматривать как выражение закона изменения во времени τ числа ожидающих. Если в начальный момент $\tau=0$ число ожидающих составляло $NP(>0)$, то $NP(>\tau)$ представит число оставшихся на ожидании по истечении времени τ . Поэтому выражение

$$N \int_0^\infty P(>\tau) d\tau$$

представит собой общую сумму времени ожидания всеми абонентами. Сле-



Фиг. 1

довательно, среднее время ожидания γ , определяемое как общая сумма времени ожидания, отнесенная к числу ожидающих N , выразится формулой

$$\gamma = \int_0^{\infty} P(>\tau) d\tau = P(>0) \frac{\theta}{s-y} \quad (2.1)$$

При $\theta = t^*$ отсюда получаем формулу Эрланга и Молина для экспоненциального распределения t . При $s=1$ имеем $P(>0)=y$ и из (2.1) получается формула, найденная А. Я. Хинчином для случая одной линии при любом распределении t . При $\theta = \frac{1}{2} t^*$, т. е. для постоянного времени занятия, из (1.11) и (2.1) имеем

$$P(>\tau) = P(>0) e^{-2(s-y)\tau/t^*}, \quad \gamma = P(>0) \frac{t^*}{2(s-y)} \quad (2.2)$$

Эти решения по форме далеки от решений, полученных Кроммелином и Полячеком для данного частного случая. Тем не менее результаты вычислений по первой формуле (2.2) чрезвычайно близко совпадают с вычислениями названных авторов уже при $s=5$.

Для одних ожидающих среднее время ожидания γ_0 согласно (2.1) будет

$$\gamma_0 = \frac{\gamma}{P(>0)} = \frac{\theta}{s-y} \quad (2.3)$$

3. Для расчетных целей могут быть использованы nomogramмы, подобные приведенной на фиг. 1. Здесь представлено соотношение между трафиком y (по оси ординат), временем ожидания τ/θ (по оси абсцисс) и необходимым числом линий s (отдельные кривые) при заданной вероятности $P(>\tau)=0.01$.

Кривые, изображенные сплошными линиями, построены для случая экспоненциального распределения времени занятия, когда $V=1$ и $\theta=t^*$. Таким образом, в основу этих кривых положена формула Эрланга-Молина. Однако ими можно пользоваться при любом распределении времени занятия t . Для этого достаточно определить из (1.7) соответствующее значение θ .

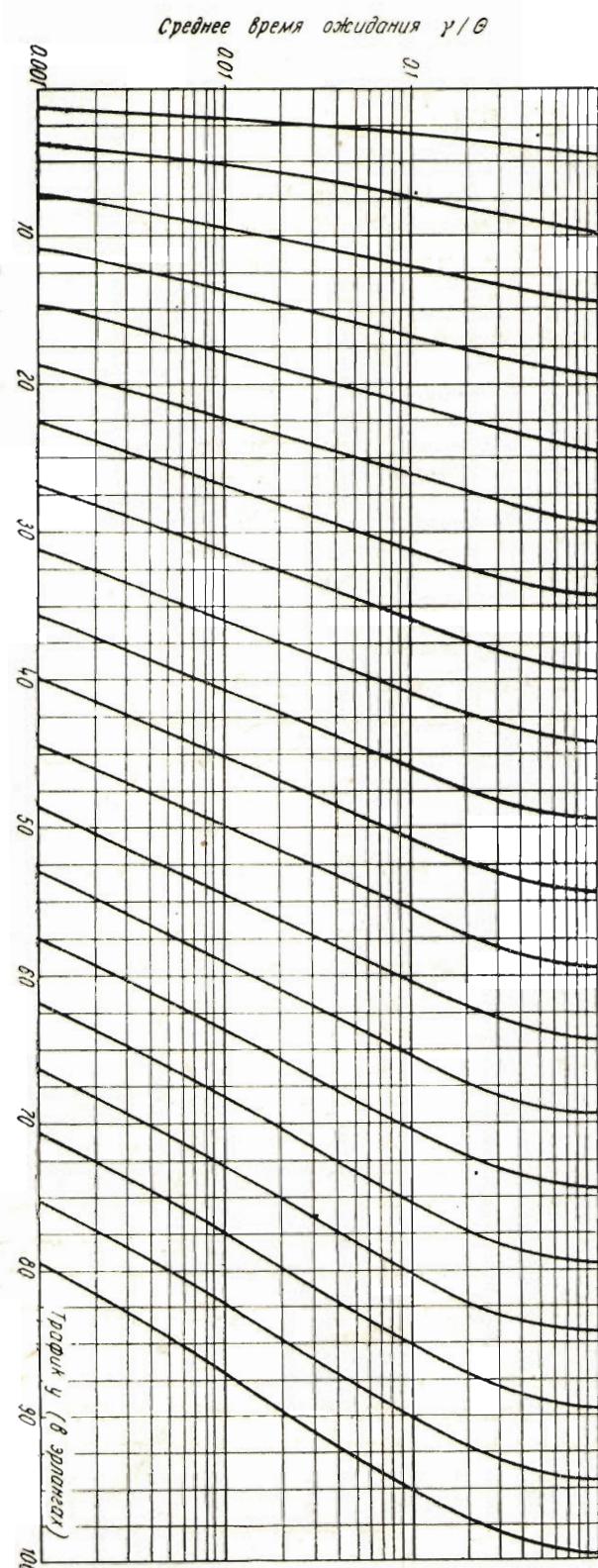
Использование nomogramm фиг. 1 покажем на примере.

Пусть трафик $y=40$ эрлангов, и мы задаемся условием, чтобы время ожидания τ не превосходило 0.75 времени занятия, т. е. $0.75 t^*$ и, следовательно, отношение τ/t^* не превосходило 0.75 с вероятностью 0.01. Тогда при экспоненциальном распределении времени занятия, когда $t^*=\theta$, мы пользуемся фиг. 1 и определяем по ней для значения $\tau/\theta=0.75$ необходимое число линий $s=45$. Сокращая допускаемое время ожидания, например, в пять раз, определяем тем же путем при $\tau/\theta=0.15$ число линий $s=51$.

При постоянстве времени занятия, когда $V=0$, и $\theta=\frac{1}{2}t^*$, мы берем для прежних же условий по оси абсцисс значение $\tau/\theta=2\tau/t^*=1.5$ и определяем при трафике $y=40$ необходимое число линий $s=43$. При сокращении τ/θ попрежнему в пять раз имеем для $\tau/\theta=0.3$ число линий $s=48$.

При большом времени допускаемого ожидания различие в числе необхо-

¹ Трафик принято выражать общим временем занятия всех линий данной группы всеми абонентами на протяжении одного часа. Единице измерения трафика, определяемой как часо-занятие за один час, на Международной конференции по электросвязи, происходившей в 1946 г. в Монтере (Швейцария), присвоено наименование эрланг.



Фиг. 2.

димых линий при различных законах распределения t оказывается мало существенным. При малой величине допускаемого времени ожидания τ/θ порядка $0.2 \div 0.4$ это различие становится значительным.

Пунктирные линии на фиг. 1 изображают соотношения между трафиком, временем ожидания и числом линий для случая постоянства времени занятия t , каждая точка пунктирной кривой делит пополам абсциссу соответствующей точки сплошной кривой.

Сравнение двух серий кривых показывает практические границы изменения числа линий при различных законах распределения t . Это сравнение подтверждает значительное различие необходимого числа линий при сравнительно малом времени ожидания (но не слишком близком к нулю) и уменьшение этого различия при увеличении допускаемого времени ожидания.

Определяемое по (1.11) выражение распределения вероятности ожидания выше τ в виде экспоненциальной кривой позволяет установить строгое соответствие между двумя основными характеристиками качества обслуживания — средним временем ожидания γ и долей случаев ожидания выше времени τ , т. е. $P(>\tau)$.

Это соответствие позволяет при расчетах исходить из среднего времени ожидания

γ , определяемого формулой (2.1). Соответствующая номограмма представлена фиг. 2.

4. В предшествующем исследовании предполагалось, что число клиентов данного вида обслуживания (абонентов) неограниченно велико и поэтому вероятность возникновения вызова за время Δt , определяемое (1.2), не зависит от числа уже занятых абонентов (переговором и ожиданием освобождения линий). При ограниченном числе абонентов N такое допущение, вообще говоря, неправильно. Если принять, что вероятность произвести вызов на протяжении заданного промежутка времени одинакова для каждого абонента, то общая величина вероятности вызова пропорциональна числу незанятых абонентов, т. е. изменяется по сравнению с (1.2) в отношении $(N-n)/(N-y-T)$ при n занятых линиях, причем $n < s$, и в отношении $(N-s-m)/(N-y-T)$ при занятости всех s линий и при m ожидающих обслуживания абонентах¹. Через T обозначена общая сумма времени ожидания всеми абонентами за 1 час (трафик ожидания, выраженный в часо-абонентах за один час). Таким образом

$$\begin{aligned} p'^+ &= y \frac{N-n}{N-y-T} \frac{\Delta t}{t^*} = (N-n) v \frac{\Delta t}{t^*} && \text{при } n \leq s \\ p'^+ &= y \frac{N-s-m}{N-y-T} \frac{\Delta t}{t^*} = (N-s-m) v \frac{\Delta t}{t^*} && \text{при } m > 0 \end{aligned} \quad (4.1)$$

Здесь через v обозначена величина трафика по расчету на одного незанятого абонента (ни переговором, ни ожиданием):

$$v = \frac{y}{N-y-T} \quad (4.2)$$

Используя значения элементарных вероятностей по (4.1) и (1.4), можем получить тем же методом, что и ранее, решения для случая ограниченного числа абонентов в следующих формах²:

а) для вероятностей $P(>\tau)$ и $P(>0)$

$$P(>\tau) = P(>0) \frac{S_{m\tau}}{S_m}, \quad P(>0) = \frac{S_m}{S_n + S_m} \quad (4.3)$$

где

$$S_{m\tau} = C_{N-1}^s v^s \sum_{v=0}^{N-s-1} C_{N-s-1}^v v! \left(\frac{v}{s}\right)^v \sum_{k=0}^v \left(\frac{s\tau}{\theta}\right)^k \frac{1}{k!} e^{-s\tau/\theta} \quad (4.4)$$

$$S_m = C_{N-1}^s v^s \sum_{v=0}^{N-s-1} C_{N-s-1}^v v! \left(\frac{v}{s}\right)^v, \quad S_n = \sum_{v=0}^{s-1} C_{N-1}^v v^v \quad (4.5)$$

В практических расчетах по этим формулам v используется как вспомогательная переменная. Задаваясь значениями v , а также значениями срока ожидания τ , можно найти соответствующие значения $P(>0)$ и $P(>\tau)$.

¹ Множитель $1/(N-y-T)$ определяется условием, что

$$\sum_{v=0}^{s-1} \frac{N-v}{N-y-T} P_v + \sum_{v=0}^{\infty} \frac{N-s-v}{N-y-T} W_v = 1$$

Фрай^[8] применяет при аналогичном выводе множитель $1/(N-y)$, что приводит к неточности результата тем большей, чем больше сумма времени ожидания T .

² Подробности см. [10].

Для того чтобы связать $P(>\tau)$ с величиной трафика y , имеем по (4.2)

$$y = \frac{(N-T)v}{v-1} \quad (4.6)$$

Трафик ожидания T определяется формулой

$$T = \frac{NS_T}{S_T + (S_n + S_m)(v+1)/v} \quad (4.7)$$

где

$$S_T = \frac{1}{s} C_{N-1}^s v^s \sum_{v=0}^{N-s-1} C_{N-s-1}^v (v+1)! \left(\frac{v}{s}\right)^v \quad (4.8)$$

Нетрудно показать, что система решений, выражаемая равенствами (4.3)–(4.8), при $N \rightarrow \infty$ приводится к (1.11). Можно убедиться также, что при $y \rightarrow 0$ имеем согласно (4.3), что $P(>0) \rightarrow 0$ и $P(>\tau) \rightarrow 0$.

При $y \rightarrow s$ имеем $P(>\tau) \rightarrow 1$. Наконец, при $s=N$ имеем $P(>\tau)=0$ для любых значений y и τ .

б) Для среднего времени ожидания γ и γ_0 при ограниченном числе абонентов и при любом распределении t можно получить

$$\gamma = P(>0) \theta \frac{S_T}{S_m} \quad (4.9)$$

Для одних ожидающих

$$\gamma_0 = \theta \frac{S_T}{S_m}$$

При $N \rightarrow \infty$ эти решения совпадают с (2.1) и (2.3).

Поступила в редакцию

14 IX 1946

F. N. BUCHMAN.—THE PROBLEM OF WAITING TIME

The paper investigates the relationship between the number of instruments, the average level of non-uniform loading and waiting time of clients, in the work of public services (telephones, etc.). Solutions which might have been applicable to the present case have heretofore been given only for particular cases. The present paper gives a general form of the asymptotic solution for any distribution of the working time of the instruments t , and for both limited and unlimited numbers of clients N . For $N \rightarrow \infty$ which is of the most considerable practical interest diagrams are given showing the number of instruments necessary for a given waiting time probability, as well as for a given average waiting time.

ЛИТЕРАТУРА

1. Erlang A. K. Elektroteknikeren. 1917. Jan. (Electrotechnische Zeitschrift. 1918. Nr. 51).
2. Molina E. C. Bell System Technical Journal. 1927. Vol. 6. N. 3.
3. Колмогоров А. Н. Математический сборник. 1930, № 38.
4. Ekelöf S. Ericsson Review. 1930. N 7—9.
5. Stommel C. D. Post Office Electrical Engineers' Journal. 1932. Vol. 25. vol. 26.
6. Хинчин А. Я. Математический сборник. 1932. № 39. Книга 4.
7. Pollaczek F. Matematische Zeitschrift. 1930. Nr. 32; 1934. Nr. 38.
8. Фрай Т. Теория вероятностей для инженеров. Гостехтеориздат. 1934.
9. Бухман Е. Н. Вестник статистики. 1927. № 3; 1928. № 1.
10. Бухман Е. Н., Подгородецкий П. А. Статистика связи. Связьиздат. 1947.