Лабораторная работа №2
Расчет основных сетевых характеристик

1. Основные вопросы теории телетрафика

В повседневной жизни нам приходится постоянно иметь дело с обслуживанием, т.е. удовлетворением некоторых потребностей, и очень часто с очередями, когда обслуживание является массовым. Примерами процессов массового обслуживания могут служить продажа билетов в железнодорожных, театральных и других кассах, обслуживание бригадой рабочих группы станков, осуществление телефонной связи и т.д. Естественно, что во всех случаях большое значение имеет степень удовлетворения потребности в обслуживании, или качество обслуживания. Так, при осуществлении телефонной связи важно знать, как долго придется ожидать соединения с требуемым абонентом после заказа междугородного разговора при ручном способе установления соединений или сколько в среднем попыток необходимо сделать для установления соединения при автоматическом способе.

Количественная сторона процессов массового обслуживания является предметом раздела прикладной математики, которую советский математик А.Я. Хинчин (1894–1959 гг.) назвал теорией массового обслуживания. Родилась теория массового обслуживания в первой четверти XX в. вследствие возникновения потребностей разработки математических методов для оценки качества функционирования телефонных систем. Основоположником теории телетрафика, из которой выросла теория массового обслуживания, является датский ученый А.К. Эрланг (1878–1929 гг.) – сотрудник Копенгагенской телефонной компании.

В теории массового обслуживания все рассматриваемые объекты объединяются под общим названием *системы массового обслуживания*. Одним из классов систем массового обслуживания являются системы распределения информации (системы телетрафика).

Системой распределения информации могут быть совокупность коммутационных приборов, часть или весь коммутационный узел либо сеть связи, которые обслуживают по определенному алгоритму телефонные, телеграфные и другие сообщения.

В настоящее время методы теории массового обслуживания используются для решения самого широкого круга задач – от бытового обслуживания до космических исследований, однако определяющую роль в развитии теории массового обслуживания продолжает играть одна из ее ветвей – теория телетрафика.

Предметом теории телетрафика является количественная сторона процессов обслуживания потоков сообщений в системах распределения информации.

Основная цель теории телетрафика заключается в разработке методов оценки качества функционирования систем распределения информации. В соответствии с этим на первом месте в теории телетрафика стоят задачи анализа, т.е. отыскание зависимостей и значений величин, характеризующих качество обслуживания, от характеристик и параметров входящего потока вызовов, схемы и дисциплины обслуживания. Эти задачи в начальный период развития телефонной техники были более актуальными, чем задачи синтеза, и решались, как правило, с помощью теории вероятностей, поэтому наиболее значительные результаты на сегодняшний день получены при решении задач анализа.

Развитие координатной и особенно квазиэлектронной и электронной коммутационной техники поставило перед теорией телетрафика сложные вероятностно-комбинаторные задачи синтеза, в которых требуется определить структурные параметры коммутационных систем при заданных потоках, дисциплине и качестве обслуживания.

Близкими к задачам анализа и синтеза являются задачи оптимизации. Эти задачи при проектировании систем распределения информации формулируются следующим образом: определить такие значения структурных параметров коммутационной системы (алгоритмы функционирования), для которых:

1) при заданных потоках, качестве и дисциплине обслуживания стоимость или объем оборудования системы распределения информации минимальны;

2) при заданных потоках, дисциплине обслуживания и стоимости качественные показатели функционирования системы распределения информации оптимальны.

При эксплуатации систем распределения информации задача оптимизации формулируется как задача управления потоками вызовов или структурой системы для достижения наилучших показателей качества функционирования. Из-за больших вычислительных трудностей задачи оптимизации систем распределения информации начали ставить и решать в последние два десятилетия после появления быстродействующих ЭВМ.

Основным математическим аппаратом теории телетрафика являются теория вероятностей, математическая статистика и комбинаторика.

Значительные результаты теории телетрафика получены благодаря сформулированному А.К. Эрлангом понятию статистического равновесия: *вероятностный процесс находится в состоянии статистического равновесия, если его вероятностные характеристики не зависят от времени*. Понятие статистического равновесия не только стимулировало развитие теории телетрафика, но также способствовало практическому применению и дальнейшему развитию теории вероятностей.

Методы математической статистики применяют при оценке результатов наблюдений за параметрами потоков вызовов и показателями качества обслуживания в действующих системах распределения информации, а также при моделировании таких систем.

При анализе, синтезе и оптимизации структурно-сложных систем распределения информации кроме вероятностных методов используют комбинаторные и алгебраические методы, теорию множеств, принципы системного подхода (системотехники). Основными методами решения задач в теории телетрафика являются аналитические, численные и метод статистического моделирования.

Аналитические методы позволяют решать задачи теории телетрафика в тех случаях, когда структура системы, характеристики потока и дисциплина обслуживания относительно просты. При этом рассматриваются все возможные состояния системы, определяемые положением каждой точки коммутации или другого элемента системы при наиболее подробном ее описании. Такие состояния называются микросостояниями системы. Каждый раз, когда поступает новый вызов, заканчивается какая-либо фаза работы управляющего устройства по установлению соединения или заканчивается соединение, система меняет свое микросостояние. Для каждого микросостояния записывается уравнение статистического равновесия. Решая систему таких уравнений, находят точное решение задачи в пределах принятой модели.

Для сложных систем число микросостояний так велико, что решить систему уравнений статистического равновесия не представляется возможным даже с помощью самых быстродействующих ЭВМ. Более перспективным является так называемый макроподход. В сложной системе с очень большим числом микросостояний имеется тот или иной признак, по которому микросостояния объединяются в классы-макросостояния. Путем усреднения определяются интенсивности переходов из одних макросостояний в другие. Для каждого макросостояния записывается уравнение статистического равновесия. В результате решения системы таких уравнений выводятся точные или приближенные формулы для вероятностей макросостояний. Чтобы представить трудности, связанные с использованием аналитических методов, достаточно указать, что число микросостояний неполнодоступного пучка из v линий оценивается как 2^{υ}. Например, при $\upsilon = 20$ число состояний более 10^6. Для решения задач такой размерности с помощью ЭВМ используют специальные алгоритмы, позволяющие находить приближенные решения итерационными или другими численными методами.

Наиболее универсальным методом, который пригоден для решения задач практически любой сложности, является метод статистического моделирования. Метод заключается в построении математической модели системы, реализация которой осуществляется в виде программы для ЭВМ. Моделирование позволяет получить численные результаты, характеризующие качество обслуживания при заданных параметрах потока, схемы и дисциплины обслуживания. Однако в силу специфики этот метод менее удобен по сравнению с аналитическим и численным методами при определении скрытых закономерностей функционирования или зависимостей между отдельными характеристиками системы. Наиболее универсальный метод решения сложных задач – метод статистического моделирования. Во многих случаях разумное сочетание аналитических и численных методов с методом статистического моделирования позволяет детально проанализировать исследуемую систему.

При малых значениях параметров системы удается получить решение с помощью точных аналитических методов и проанализировать предельные случаи при асимптотическом поведении характеристик изучаемой системы. Полученные сведения дополняют результатами статистического моделирования в области реальных значений параметров системы.

Оценивая результаты исследований систем распределения информации любыми математическими методами, следует помнить, что математика оперирует не реальными системами, а их математическими моделями. Так как математические модели всегда лишь приближенно описывают реальные системы, то никакие математические методы не могут заменить исследований, проводимых на реально функционирующих системах.

2 ПОТОКИ ВЫЗОВОВ

Потоком вызовов (в общем случае – событий) называется последовательность вызовов, поступающих через какие-либо интервалы или в какие-либо моменты времени. В теории массового обслуживания под потоком вызовов принято понимать не только последовательность вызовов, поступающих от группы абонентов или группы устройств телефонной сети, но и другие последовательности событий, например, поток телеграмм, поток писем, поток неисправностей отдельных коммутационных устройств или телефонных сооружений в целом, поток информации, поступающей на ЭВМ, поток неисправностей в станках и т.п. Рассматриваемые в настоящей главе свойства, характеристики, закономерности потоков вызовов не ограничиваются узкими рамками изучения потоков телефонных вызовов, а имеют более широкую область применения.

Следует различать детерминированный и случайный потоки вызовов. Детерминированный поток вызовов – это последовательность вызовов, в которой вызовы поступают в определенные, строго фиксированные неслучайные моменты или через определенные, строго фиксированные, неслучайные промежутки времени. Случайный поток вызовов отличается от детерминированного тем и только тем, что моменты поступления вызовов и промежутки времени между вызовами являются не строго фиксированными, а случайными величинами.

Детерминированные потоки являются частным случаем случайных потоков и на практике встречаются редко. Примерами их могут служить: поток сеансов связи с искусственными спутниками Земли, поток поступления деталей и выхода изделий ритмично работающего завода и т.п. Строго говоря, даже в таких потоках часто имеют место случайности. В связи с этим в теории телетрафика основное внимание уделяется рассмотрению случайных потоков вызовов.

Условимся в дальнейшем случайные величины обозначать прописными (большими) буквами, а их возможные значения – соответствующими строчными (малыми) буквами.

Поток вызовов может быть определен тремя эквивалентными способами: последовательностью вызывающих моментов t_1, t_2,...,t_n, последовательностью промежутков времени между вызывающими моментами z_1, z_2,...,z_n и последовательностью чисел k_1, k_2,...,k_n, определяющих количество вызовов, поступающих в течение заданных отрезков времени $[t_0, t_1)$, $[t_0, t_2)$,..., $[t_0, t_n)$. При этом под вызывающим моментом понимается момент одновременного поступления одного, двух и более вызовов; для вызывающих моментов всегда, если $t_i > t_i - 1$, то $z_i > 0$, в то время как для момента поступления вызова $t_i. \geq t_i - 1$ и $z_i \geq 0$.

Определение случайного потока вызовов связано с определением в вероятностном смысле либо последовательности вызывающих моментов, либо последовательности промежутков между вызывающими моментами, либо последовательности чисел вызовов, поступающих в течение отрезков времени $[t_0, t_1)$, $[t_0, t_2)$,...,$[t_0, t_n)$.

Для задания случайных потоков вызовов, как и любых других случайных величин и процессов, используют функции распределения. Функцией распределения вероятностей некоторой случайной величины X называется функция

$$F(x) = P\{X < x\}, \tag{2.1}$$

определяющая вероятность того, что $X < x$, где x – определенная, заданная величина.

С учетом вышеизложенного для задания случайного потока вызовов могут быть использованы следующие эквивалентные способы:

1) совместный закон распределения n случайных вызывающих моментов

$$P = \{T_i < t_i, i = 1,2,...,n\} = P\{T_1 < t_1, T_2 < t_2,...,T_n < t_n\}, \tag{2.2}$$

где T_i – i-й вызывающий момент; n может принимать любое значение;

2) совместный закон распределения n случайных промежутков времени между вызывающими моментами

$$P = \{Z_i < z_i, i = 1,2,...,n\} = P\{Z_1 < z_1, Z_2 < z_2,...,Z_n < z_n\}, \tag{2.3}$$

где Z_i – промежуток времени между $(i - 1)$- и i-м вызывающими моментами; n может принимать любое значение;

3) совместный закон распределения числа вызовов K на n отрезках времени $[t_0, t_1)$, $[t_0, t_2)$, ..., $[t_0, t_n)$:

$$P = \{K(t_0,t_i) = k_i, i = 1,2,...,n\} = P\{K(t_0,t_1) = k_1, K(t_0,t_2) = k_2,...,K(t_0,t_n) = k_n\}, \tag{2.4}$$

где n может принимать любое значение; $k_1 \le k_2 \le .... \le k_n$; $t_1 < t_2 < ... < t_n$.

Введем некоторые ограничения на рассматриваемые случайные потоки вызовов.

Потоки вызовов подразделяются на неоднородные и однородные. В неоднородном потоке вызовов каждый вызов имеет две и более характеристики. Например, вызовы, поступающие от абонентов телефонной сети, определяются моментами их поступления, направлениями установления соединений, длительностью их обслуживания и другими характеристиками.

Аналогично телеграммы, поступающие на телеграф, характеризуются моментами их поступления, направлениями их передачи и т.д.

Однородный поток вызовов характеризуется последовательностью, определяющей только закономерность поступления вызовов, т.е. последовательностью моментов поступления вызовов или промежутков времени между вызовами, либо иным способом задания потока вызовов.

На практике потоки вызовов, как правило, являются неоднородными. Несмотря на это, целесообразно отдельно от других характеристик потоков вызовов изучить последовательности моментов поступления вызовов, поэтому в дальнейшем под потоком вызовов будем понимать однородный поток вызовов.

Ограничимся рассмотрением потоков, в которых на любой конечный отрезок времени поступает конечное число вызовов и математическое ожидание числа поступающих вызовов также является конечной величиной. Такие потоки называются финитными.

Математическое ожидание числа вызовов, поступающих в интервале времени $[0, t)$, называется ведущей функцией потока. Обозначим эту функцию $\Delta(0, t)$. Функция $\Delta(0, t)$ – неотрицательная, неубывающая и в практических задачах принимает конечное значение.

Потоки с непрерывной ведущей функцией называются регулярными, а со ступенчатой – сингулярными. Вероятность поступления хотя бы одного вызова в определенный момент времени для регулярного потока равна нулю, а для сингулярного потока в моменты разрыва ведущей функции отлична от нуля. Нас интересуют только потоки вызовов с непрерывной ведущей функцией, т.е. регулярные потоки.

Таким образом, в дальнейшем рассматриваются случайные однородные финитные регулярные потоки.

Потоки вызовов классифицируются с точки зрения стационарности, ординарности и последействия.

Стационарность потока. Поток вызовов является стационарным, если при любом n совместный закон распределения числа вызовов за промежутки времени $[t_0, t_1)$, $[t_0, t_2)$, ..., $[t_0, t_n)$

$$P = \{K(t_0, t_i) = k_i, i = 1,2,...,n\} \qquad (2.5)$$

зависит только от длины промежутков времени и не зависит от момента t_0. Иными словами, независимо от того, где на оси времени расположен промежуток времени $[t_0, t_1)$, вероятность поступления $K(t_0, t_t)$ вызовов одна и та же. Это значит, что для стационарного потока вероятность поступления некоторого числа вызовов за какой-то промежуток времени зависит от длины этого промежутка и не зависит от его начала. В противном случае поток является нестационарным.

Интенсивности потоков вызовов на телефонных сетях резко колеблются в зависимости от времени суток: количество вызовов за единицу времени в определенные дневные и вечерние часы достигает максимальной величины, а в ночные часы уменьшается почти до нуля. Это значит, что вероятность поступления какого-либо числа вызовов в определенный промежуток времени зависит от местонахождения на оси времени этого промежутка и, следовательно, поток поступающих в течение суток вызовов от любой абонентской группы на телефонную станцию является нестационарным. Заметим, что внутри ограниченного отрезка суток, например часа, нестационарность телефонного потока вызовов малоощутима, что позволяет для практических задач полагать стационарным поток телефонных вызовов, поступающих от большой абонентской группы (100 и более абонентов) за небольшой отрезок суток, исчисляемый одним-тремя часами.

Ординарность потока. Обозначим через $\pi k(t,\ t+\tau)$ вероятность поступления k и более вызовов за промежуток $[t,\ t+\tau)$. Поток вызовов является ординарным, если при $\tau \to 0$

$$\lim_{\tau\to 0} \frac{\pi_2(t,t+\tau)}{t} = 0, \tag{2.6}$$

т.е. $\pi_2(t,\ t+\tau) = o(\tau)$, где $o(\tau)$ – величина более высокого порядка малости по отношению к τ.

Ординарность потока выражает практическую невозможность одновременного поступления двух и более вызовов в любой момент времени t. Примером ординарного потока является поток вызовов, поступающий на телефонную станцию от абонентской группы любой емкости. Потоки телефонных вызовов к абонентам диспетчерской или конференц-связи на несколько адресов являются неординарными.

Последействие потока. Поток вызовов является потоком без последействия, если вероятность поступления $K(t_0,\ t_i)$ вызовов за промежутки $[t_0,\ t_i)$, $i = 1,\ 2,\ ...,\ n$

$$P\{K(0,t_i) - K(0,t_0)\} = K(t_0,t_i), i = 1,2,..,n \tag{2.7}$$

не зависит от вероятностного процесса поступления вызовов до момента t_0. Иными словами, отсутствие последействия потока означает независимость течения случайного потока вызовов после какого-либо момента времени от его течения до этого момента.

Примером потока без последействия может служить поток телефонных вызовов, поступающих от большой группы источников. Действительно, лишь небольшая часть (10–20 %) абонентской группы одновременно участвует в телефонных соединениях, поэтому вероятность поступления какого-либо числа вызовов от большой группы источников на любом отрезке времени практически не зависит от процесса поступления вызовов до начала данного отрезка. Заметим, что эта вероятность, как и вероятность (2.7), может зависеть от момента t_0 начала этого отрезка времени. Так, различные значения принимает вероятность поступления некоторого числа телефонных вызовов за равные промежутки времени в различные часы суток в силу нестационарности потока телефонных вызовов в течение суток. Поток вызовов является потоком с последействием, если вероятность поступления того или иного числа вызовов за некоторый промежуток времени зависит от процесса поступления вызовов до начала этого промежутка. Потоки вызовов от спаренных телефонных аппаратов, от малых абонентских групп, в направлениях коммутационной системы, не обеспечивающих удовлетворительного качества обслуживания абонентов телефонной связью, к интенсивно загруженным абонентам являются потоками с последействием.

К основным характеристикам потока вызовов следует отнести ведущую функцию потока, его параметр и интенсивность.

Под параметром потока $\lambda(t)$ в момент времени t понимается предел отношения вероятности поступления хотя бы одного вызова за время $[t, t + \tau)$ к длине этого отрезка времени τ при $\tau \to 0$:

$$\operatorname*{Lim}_{\tau\to 0} \frac{\pi_i(t, t+\tau)}{t} = \lambda(t), \tag{2.8}$$

т.е. параметр потока есть плотность вероятности наступления вызывающего момента в момент t. Исходя из (2.8), находим вероятность поступления одного и более вызовов за время $[t, t + \tau)$:

$$\pi(t, t+\tau) = \lambda(t)\tau + o(\tau), \ \tau \to 0\,. \tag{2.9}$$

Согласно определению стационарного потока вероятность поступления определенного числа вызовов за некоторый промежуток времени одна и та же и не зависит от месторасположения на оси времени этого промежутка. Следовательно, и плотность вероятности поступления вызовов стационарного потока, т.е. его параметр $\lambda(t)$, есть величина постоянная, не зависящая от момента, т.е. $\lambda(t) = \lambda$. Отсюда для стационарных потоков

$$\pi(t, t+\tau) = \lambda\tau + o(\tau), \ \tau \to 0\,. \tag{2.10}$$

В отличие от ведущей функции потока $\Lambda(0, t)$, определяющей математическое ожидание числа вызовов, поступающих в промежутке времени $[0, t)$, параметр потока $\lambda(t)$ характеризует не поток вызовов, а поток вызывающих моментов, и эта характеристика относится не ко всему отрезку $[0, t)$, а лишь к фиксированному моменту t.

Интенсивностью стационарного потока μ называется математическое ожидание числа вызовов, поступающих в единицу времени. Единица времени может быть выбрана произвольно, однако в теории телетрафика в качестве такой единицы большей частью принимают среднюю длительность одного занятия. Вследствие аддитивности математического ожидания для стационарного потока ведущая функция за промежуток времени $[0, t)$ равна $\Lambda(0, t) = \mu t$.

Для нестационарных потоков используются понятия средней и мгновенной интенсивностей. Средняя интенсивность потока на отрезке времени $[t1, t2)$ есть

$$\bar{\mu}(t_1, t_2) = [\Lambda(0, t_2) - \Lambda(0, t_1)]/(t_2 - t_1), \tag{2.11}$$

а мгновенная интенсивность потока в момент t

$$\mu(t) = \operatorname*{Lim}_{\tau \to 0} \frac{\Lambda(0, t+\tau) - \Lambda(0,t)}{t}. \tag{2.12}$$

Согласно определению (2.12) мгновенная интенсивность потока представляет собой производную ведущей функции потока. Так же как и параметр потока $\lambda(t)$, мгновенная интенсивность потока $\mu(t)$ относится не к отрезку времени поступления вызовов, а только к моменту t. В то же время в отличие от параметра потока, характеризующего поток вызывающих моментов, мгновенная интенсивность потока характеризует поток поступления вызовов.

Для любых потоков вызовов $\mu(t) \geq \lambda(t)$, причем для ординарных потоков $\mu(t) = \lambda(t)$. Для стационарных потоков интенсивность и параметр постоянны: $\mu(t) = \mu$, $\lambda(t) = \lambda$. Следовательно, для любых стационарных потоков $\mu \geq \lambda$, а для стационарных ординарных $\mu = \lambda$.

Классификацию потоков удобно осуществлять, принимая за основной признак последействие потока. С точки зрения последействия различают три класса потоков: без последействия, с простым последействием и с ограниченным последействием.

Начнем рассмотрение этих классов с потоков без последействия. К этому классу относятся: стационарный ординарный поток, называемый простейшим (его также называют стационарным пуассоновским), нестационарный ординарный поток, называемый нестационарным пуассоновским, и стационарный неординарный поток, называемый неординарным пуассоновским.

Определение. Простейшим потоком называется стационарный ординарный поток без последействия. Простейший поток вызовов является наиболее распространенной моделью реального потока вызовов, применяемой в системах массового обслуживания, в том числе в теории телетрафика. Действительно, как отмечалось при рассмотрении принципов классификации потоков вызовов, поток телефонных вызовов от большой группы абонентов характеризуется отсутствием последействия. Его можно считать ординарным, а при ограничении исследуемого промежутка времени 1–3 ч и стационарным. Аналогичные случайные потоки событий характерны для многих отраслей народного хозяйства.

Математическая модель простейшего потока. Определим вероятности поступления точно $k (k = 0, 1, 2, ...)$ вызовов на отрезке времени $[t_0, t_0 + t)$: $p\mathrm{k}(t_0, t_0 + t)$. Исследования будем проводить на отрезке времени $[t_0, t_0 + t + \tau)$, который можно представить состоящим из двух примыкающих друг к другу отрезков: $[t_0, t_0 + t + \tau) = [t_0, + t_0 + t) + [t, t + \tau)$.

Для того чтобы в течение отрезка $[t_0, t_0 + t + \tau)$ поступило точно k вызовов, необходимо, чтобы за первый промежуток времени $[t_0, t_0 + t)$ поступило k, или k–1, ..., или k–i, ..., или 0 вызовов и соответственно за второй промежуток 0, или 1, ..., или i, ..., или k вызовов.

Введем обозначения: $p_k(t_0,\ t_0 + t + \tau)$ n^-вероятность поступления точно k вызовов за отрезок времени $[t_0,\ t_0 + t + \tau)$; $p_{k-i}(t_0,\ t_0 + t)$ – вероятность поступления точно $k–i$ вызовов за первый отрезок времени $[t_0,\ t_0 + t)$; $p_i(t,\ t + \tau)$ – вероятность поступления точно i вызовов за второй отрезок времени $[t,\ t + \tau)$. Согласно определению простейший поток является стационарным.

Из этого следует, что вероятности поступления того или иного числа вызовов за отрезки времени $[t_0,\ t_0 + t + \tau)$, $[t_0,\ t_0 + t)$, $[t,\ t + \tau)$ не зависят от моментов начала отсчета времени, а зависят только от длины отрезков времени. Поэтому упростим обозначения как отрезков времени, так и вероятностей: $[t_0,\ t_0 + t + \tau)$ будем обозначать $[t + \tau)$; $[t_0,\ t_0 + t)$ – $[t)$; $[t,\ t + \tau)$ – $[\tau)$ и соответственно $p_k(t_0,\ t_0 + t + \tau)$ – $p_k(t + \tau)$; $p_{k-i}(t_0,\ t_0 + t)$ – $p_{k-i}(t)$; $p_i(t,\ t + \tau)$ – $p_i(\tau)$.

Простейший поток является потоком без последействия, поэтому независимыми являются события, заключающиеся в поступлении какого-либо числа вызовов за первый и второй промежутки времени, и вероятность поступления точно k вызовов за время $[t + \tau)$ для каждой реализации $i = 0, 1, ..., k$ составляет $pk(t + \tau)i = pk\text{-}i(\mathrm{t})pi(\tau)$, $i = 0, 1, ..., k$. Поскольку реализации с $i = 0, 1, ..., k$ представляют несовместимые события, то согласно формуле полной вероятности имеем

$$p_k(t+\tau) = \sum_{i=0}^{k} p_k(t+\tau)_i = \sum_{i=0}^{k} p_{k-i}(t)p_i(\tau),\ k = 1,2,... \tag{2.13}$$

Выражение (2.13) представляет собой систему, состоящую из бесконечного числа уравнений. Устремим отрезок времени τ к нулю. Вследствие ординарности простейшего потока $\pi 2(t,\ t + \tau) = o(t)$, $\tau \to 0$. Тем более вероятности поступления точно 2, 3, ... вызовов – $p_2(\tau)$, $p_3(\tau)$, ... – есть бесконечно малые более высокого порядка по отношению к τ. Следовательно, в системе уравнений (2.13) вероятности p_i имеют конечные значения только при i, равном 0 и 1.

На основании этого (2.13) преобразуются к виду

$$p_k(i+\tau) = p_{k-1}(t)p_i(\tau) + p_k(t)p_0(\tau) + o(\tau),\ k = 0,1,...,\ \tau \to 0\,. \tag{2.14}$$

Определяем вероятности $p_1(\tau)$ и $p_0(\tau)$:

$$p_1(\tau) = \pi_1(\tau) - \pi_2(\tau);\ p_0(\tau) = \pi_0(\tau) - \pi_1(\tau)\,.$$

С учетом (2.10) и (2.6)

$$p_1(\tau) = \lambda\tau + o(\tau);\ p_1(\tau) = 1 - \lambda\tau + o(\tau) \tag{2.15}$$

($\pi_0(\tau)$ – вероятность поступления 0 и более вызовов, т.е. вероятность достоверного события, она равна 1).

Подставим в систему уравнений (2.14) полученные значения вероятностей $p_1(\tau)$ и $p_0(\tau)$.

Затем, перенеся в левую часть уравнений $p_k(t)$, поделим левые и правые части уравнений на τ.

Переходя к пределу, получим

$$\frac{d}{dt}p_k(t)=\lambda p_{k-1}(t)-\lambda p_k(t),\ k=0,1,... \qquad (2.16)$$

Решив систему дифференциальных уравнений (2.16), получим формулу Пуассона

$$p_k(t)=\frac{(\lambda t)^k}{k!}e^{-\lambda t}. \qquad (2.17)$$

Таким образом, вероятность поступления точно k вызовов простейшего потока за отрезок времени t определяется формулой Пуассона. По этой причине простейший поток также называют стационарным пуассоновским потоком.

Нестационарный пуассоновский поток (который также называется потоком с переменным параметром или нестационарным простейшим потоком) есть ординарный поток без последействия, для которого в любой момент времени t существует конечный параметр $\lambda(t)$, зависящий от момента t. По аналогии с простейшим потоком в качестве математической модели нестационарного пуассоновского потока выбирается вероятность $p_k(t_0, t)$ поступления точно k вызовов за заданный промежуток времени $[t_0, \mathrm{t})$. В силу нестационарности потока эта вероятность зависит не только от длины промежутка времени $[t_0, t)$, но и от начального момента t_0:

$$p_k(t_0,t)=\frac{\left[\int\limits_{t_0}^{t}\lambda(u)du\right]^k}{k!}e^{-\int\limits_{t_0}^{t}\lambda(u)du},k=0,1,... \qquad (2.18)$$

Заметим, что для стационарного потока $\lambda(u)=\lambda$, $\int\limits_{t_0}^{t}\lambda(u)du=\lambda(i-t_0)=\lambda t(i[t_0,t)=t)$ и формула (2.18) преобразуется к (2.17).

Для неординарного пуассоновского потока, т.е. для стационарного неординарного потока без последействия, следует различать поток вызывающих моментов и поток вызовов. Поток вызывающих моментов характеризуется вероятностью появления точно i вызывающих моментов в промежутке времени t. Эта вероятность $p_i(t)$ определяется формулой Пуассона (2.17).

В каждый вызывающий момент поступает l вызовов. Величина l, называемая характеристикой неординарности потока, может быть постоянной и переменной. Если l является постоянной величиной, то с вероятностью $p_i(t)$ суммарное число вызовов, поступающих за отрезок времени t, составляет $k = li$.

Для неординарного пуассоновского потока с переменной величиной l, в котором в каждый вызывающий момент с вероятностью ωl поступает l вызовов $\left(\sum_{l=1}^{r}\omega_l = 1\right)$, также получена формула, определяющая вероятность $p_k(t)$ поступления точно k вызовов за промежуток времени t. Ввиду громоздкости эту формулу здесь не приводим. Параметр такого потока для каждого значения l равен $\lambda\omega_l$. Отсюда общий параметр потока $\left(\sum_{l=1}^{r}\lambda\omega_l = 1\right)$ такой же, как и для потока вызывающих моментов, т.е. для простейшего потока. Интенсивность μ неординарного пуассоновского потока, как и любого стационарного неординарного потока, больше его параметра λ. Действительно, $\mu = \sum_{l=1}^{r} l\lambda\omega_l = \lambda\sum_{l=1}^{r} l\omega_l > \lambda$.

Основной характеристикой потока с простым последействием является зависимость параметра потока от состояния коммутационной системы в любой момент времени t.

Коммутационная система имеет множество состояний s, которые различаются числом занятых входов, выходов и соединительных путей между входами и выходами коммутационной системы, номерами занятых входов и выходов, номерами соединительных путей между входом и выходом, числом занятых или свободных источников вызовов и т.д.

Так как коммутационная система всегда имеет конечное число входов, выходов, соединительных путей, то конечным является и число возможных состояний системы обслуживания (хотя оно может быть и очень велико). Такие состояния коммутационной системы будем называть микросостояниями. Состояния коммутационной системы, различающиеся только числом занятых входов (или выходов), называются макросостояниями.

Исследования процесса обслуживания коммутационной системой поступающего потока вызовов следует производить с учетом микросостояний системы в тех случаях, когда структура рассматриваемой коммутационной системы такова, что вероятность ее состояния в любой произвольный момент времени t зависит как от числа занятых входов (или выходов), так и от того, какие именно входы (или выходы) заняты и по каким соединительным путям осуществляются соединения между каждым входом и выходом. Если же структура рассматриваемой коммутационной системы такова, что вероятности ее состояний в любой произвольный момент t зависят только от числа занятых входов (или выходов), то исследования процесса обслуживания коммутационной системой поступающих вызовов можно производить только с учетом макросостояний системы.

Под параметром потока в состоянии $s(t)$ будем понимать предел

$$\lambda_s(t) = \mathop{\mathrm{Lim}}_{\tau \to 0} = \frac{\pi_1(t, t + \tau / s(t))}{\tau}, \qquad (2.19)$$

где $\pi_1(t,\ t + \tau/s(t))$ – вероятность поступления за промежуток $[t,\ t + \tau)$ одного и более вызовов, если в момент t коммутационная система находится в состоянии $s(t)$.

Данное определение позволяет сформулировать понятие потока с простым последействием. Под потоком с простым последействием понимается ординарный поток, для которого в любой момент времени t существует конечный параметр потока в состоянии $s(t)$ (2.19), зависящий только от состояния $s(t)$ коммутационной системы в момент t и не зависящий от процесса обслуживания вызовов до момента t.

Параметр потока с простым последействием в любой момент времени t зависит от состояния системы в этот момент времени, а состояние системы $s(t)$, в свою очередь, зависит от процесса поступления и обслуживания вызовов до момента t. Такое последействие принято называть простым, поскольку для определения параметра потока в момент t достаточно ограничиться знанием состояния системы $s(t)$ в этот момент. Поток с простым последействием является нестационарным, так как его параметр зависит от момента t. Заметим, что эта зависимость проявляется через состояние $s(t)$. Для каждого конкретного состояния параметр потока с простым последействием является постоянной величиной.

3 НАГРУЗКА, ЕЁ ИЗМЕРЕНИЕ, ПРОГНОЗИРОВАНИЕ, РАСПРЕДЕЛЕНИЕ

Важным показателем качества цифровых систем связи является объем переданных данных за время Т. При этом встает вопрос: как оценивать объем переданной информации? Характеристика объема должна быть универсальной величиной и не относиться к различным типам передаваемых сообщений, например, числу знаков, слов, битов и т.п. Наиболее удачной характеристикой является время, необходимое для передачи той или иной информации по каналам связи. Причем чем больше информации передается, тем больше затрачивается времени на ее передачу, и дольше работают обслуживающие приборы, т.е. они больше нагружены.

Если за время T поступило $C(T)$ вызовов, каждый из которых занимал канал связи в течение времени t_i, то суммарное время обработки вызовов составит величину

$$A(T) = \sum_{i=1}^{C(T)} t_i = C(T) t_{cp}, \tag{3.1}$$

где $t_{\text{ср}}$ – среднее время обслуживания одного вызова.

Величина $A(T)$ получила название *нагрузка*. Различают три вида нагрузок: поступающая нагрузка, обслуженная нагрузка и потерянная нагрузка.

Поступающая нагрузка – это время, которое необходимо затратить на обслуживание N источников вызовов:

$$A_{пост} = t_{cp} \sum_{i=1}^{N} C_i, \tag{3.2}$$

где $t_{\text{ср}}$ – среднее время обслуживания одной заявки;

c_i – число вызовов, поступивших от i-го абонента.

При анализе систем связи часто вместо понятия *нагрузка* используют понятие *интенсивность нагрузки*, которая определяется следующим образом. Интенсивность нагрузки – это среднее число вызовов, поступивших на вход системы распределения информации от N источников вызовов за время, равное средней длительности одного занятия $t_{\text{ср}}$:

$$\lambda_{\text{пост}} = N c_{\text{ср}}, \tag{3.3}$$

где $c_{\text{ср}}$ – среднее число вызовов от одного абонента за время $t_{\text{ср}}$.

Из полученного выражения следует, что величина поступающей нагрузки связана с интенсивностью соотношением

$$A_{\text{пост}} = \lambda t_{\text{ср}}. \tag{3.4}$$

Нагрузка и ее интенсивность измеряются в Эрлангах (Эрл). Интенсивность поступающей нагрузки в 1 Эрл создается потоком поступающих вызовов с интенсивностью в один вызов за среднее время занятия канала связи.

Интенсивность поступающей нагрузки на систему распределения информации удобно выражать через интенсивность заявок от одного источника вызовов, которую называют удельной интенсивностью.

Удельная интенсивность поступающей нагрузки – это среднее число вызовов, поступающих на вход системы распределения информации от одного источника за время, равное средней длительности одного занятия $t_{ср}$:

$$\alpha_{уд} = \lambda/N = c_{ср}. \tag{3.5}$$

Обслуженная нагрузка – это общее время занятия каналов связи за период времени, равный средней длительности обслуживания $t_{ср}$:

$$A_{обсл} = n_{ср} t_{ср}, \tag{3.6}$$

где $n_{ср}$ – среднее число занятых каналов связи в течение времени $t_{ср}$.

Интенсивностью обслуженной нагрузки называется среднее число занятых каналов связи за время $t_{ср}$

$$\mu_{обсл} = A_{обсл}/t_{ср} = n_{ср}. \tag{3.7}$$

Потерянная нагрузка представляет собой разность между поступающей и обслуженной нагрузками

$$A_{пот} = A_{пост} - A_{обсл}. \tag{3.8}$$

Таким образом, поступившая заявка может быть либо обслужена, либо отброшена. В соответствии с этим различают три типа систем распределения информации: без потерь, с потерями и с ожиданием.

Системы без потерь возможны только в том случае, когда число источников вызовов S меньше или равно числу каналов связи V. Очевидно, что выполнить это условие не всегда возможно и экономически неоправданно, поэтому часто в системах связи имеет место соотношение $S >> V$, в результате возникают потери $A_{пот}$. Третий тип систем позволяет избежать потерь за счет использования очереди заявок, заставших все каналы занятыми. В таких системах объем поступающей и обслуженной нагрузок одинаковый, так как поступивший вызов либо обслуживается сразу, либо встает очередь до освобождения канала связи. Таким образом, вызов не теряется, а лишь задерживается на некоторое время.

Характеристикой качества всех трех видов систем распределения информации является вероятность обслуживания заявки q:

$$q = \frac{\lambda_{\text{обсл}}}{\lambda_{\text{пост}}}. \tag{3.9}$$

или вероятность отказа в обслуживании p:

$$p = \frac{\lambda_{\text{пот}}}{\lambda_{\text{пост}}}. \tag{3.10}$$

Величина получила название *вероятность блокировки вызова*. При этом за единицу измерения качества обслуживания принята тысячная доля единицы или промилле: 1 ‰ = 0,001.

Кроме потерь по вызовам существуют еще потери по нагрузке p_x и по времени p_t. Вероятность потерь по нагрузке определяется выражением

$$p_{\text{н}} = \frac{\lambda_{\text{пот}}}{\lambda_{\text{пост}}}. \tag{3.11}$$

а по времени

$$p_t = t_{\text{ср}}/T, \tag{3.12}$$

где $t_{\text{ср}}$ – среднее время занятости каналов связи;

T – общее время работы системы распределения информации.

Для оценивания качества систем с ожиданием используется вероятность превышения времени ожидания $t_{\text{ож}}$ заданной величины t:

$$P(t_{\text{ож}} > \tau) = \beta. \tag{3.13}$$

Если величина $\beta = 0{,}1$, то система распределения информации имеет высокое качество обслуживания, при $\beta = 0{,}2$ – пониженное. Вероятность $P(t_{\text{ож}} > \tau)$ рассчитывается на основе накопленной статистики: числа вызовов, простоявших в очереди времени больше величины τ и общего числа поступивших заявок:

$$P(t_{\text{ож}} > \tau) = \frac{C(t_{\text{ож}} > \tau)}{Y_{\text{пост}}}. \tag{3.14}$$

Измерение интенсивности обслуженной нагрузки. Рассмотрим систему распределения информации, состоящую из пяти каналов. На вход этой системы за промежуток времени (t_1, t_2) поступают информационные потоки. На основе наблюдений за работой системы строят график обслуженной нагрузки $i(t)$, которая может принимать значения от 0 до 5: число занятых каналов в момент времени t (рис. 3.1). На графике нагрузки отметим длительности отдельных состояний пучка $z_i(k) = 1,2,\ldots,n$, в течение времени которых занято i линий, где n – число таких состояний на интервале (t_1, t_2).

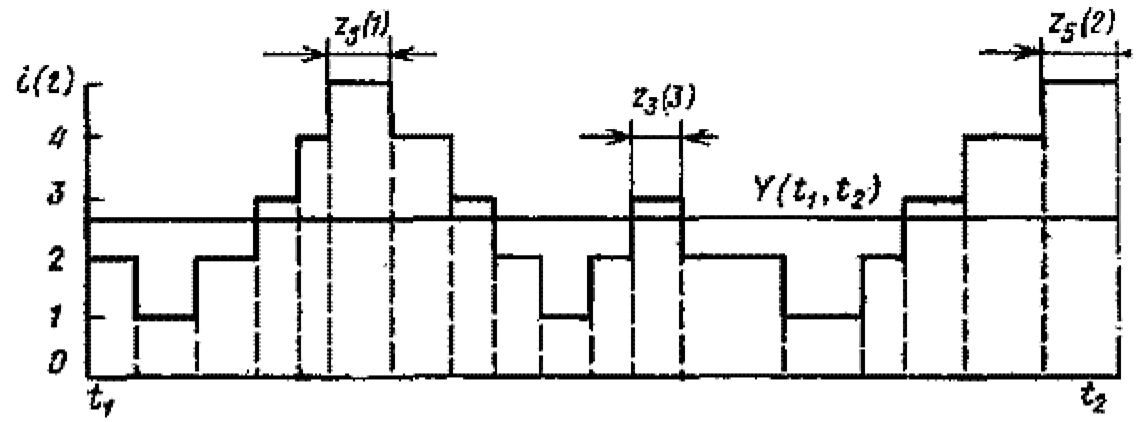


Рис. 3.1. Распределение нагрузки в пятиканальной системе

Обозначим через $z_i(t_1, t_2)$ сумму всех n промежутков времени $z_i(k)$, когда в пучке занято ровно i линий:

$$z_i(t_1,t_2)=\sum_{k=1}^{n} z_i(k)\,. \tag{3.15}$$

Очевидно, что сумма длительностей всех состояний пучка линий равна общей длительности процесса наблюдения

$$\sum_{i=0}^{V} z_i(t_1,t_2)=t_2-t_1\,, \tag{3.16}$$

где V – число каналов.

На основе наблюдений $i(t_1, t_2)$ и $z_i(t_1, t_2)$ определим работу, совершаемую системой за время наблюдения, как площадь под графиком $i(t)$:

$$A(t_1,t_2)=\sum_{i=0}^{V} i z_i(t_1,t_2)\,. \tag{3.17}$$

Из рис. 3.1 также видно, что работа есть произведение интенсивности нагрузки $\lambda(t_1, t_2)$ на величину промежутка наблюдения (t_1, t_2):

$$A(t_1, t_2) = \lambda(t_1, t_2)(t_2 - t_1). \tag{3.18}$$

Работа измеряется в Эрлангах-часах (Эрл-ч) или, как было ранее, в часозанятиях (ч.зан). Так, например, работу в 1 Эрл-ч выполняет линия, непрерывно занятая в течение одного часа; работу в 2 Эрл-ч выполняют две линии, непрерывно занятые в течение часа, либо одна линия, непрерывно занятая в течение двух часов.

Из выражения (3.18) следует, что интенсивность нагрузки можно определить по формуле

$$\lambda(t_1, t_2) = A(t_1, t_2)/(t_2 - t_1). \tag{3.19}$$

При таком подходе к измерению интенсивности нагрузки необходимо за время измерения ($t_2 - t_1$) фиксировать число занятых каналов связи в каждый момент времени t и сумму этих величин делить на время наблюдения.

Распределение интенсивности нагрузки по времени и ее измерение. Интенсивность поступающей нагрузки в системах связи, как правило, обладает резко выраженной нестационарностью. Причем колебания интенсивности нагрузки характеризуются определенной периодичностью, что необходимо учитывать при ее прогнозировании.

Для удовлетворительного качества обслуживания абонентов в любое время суток расчет объема оборудования и пропускной способности каналов связи необходимо выполнять, учитывая значения интенсивности нагрузки в тот час, когда она является наибольшей. Этот час называется часом наибольшей нагрузки (ЧНН).

ЧНН – это непрерывный интервал времени продолжительностью 60 мин, в течение которого средняя интенсивность нагрузки является наибольшей. Для интервала наибольшей интенсивности нагрузки, равного трем часам, используют понятие *период наибольшей нагрузки*. ЧНН и ПНН определяют с точность до 15 мин.

Степень концентрации нагрузки в ЧНН оценивается коэффициентом концентрации

$$k_{\text{ЧНН}} = \frac{\lambda_{\text{ЧНН}}}{\lambda_{\text{сут}}}, \tag{3.20}$$

где $\lambda_{\text{ЧНН}}$ – интенсивность нагрузки в ЧНН;

$\bar{\lambda}_{\text{сут}}$ – средняя интенсивность нагрузки за сутки.

Международный консультативный комитет по телефонии и телеграфии (МККТТ) рекомендовал проводить измерения наибольшей нагрузки в рабочие дни двух последовательных недель в месяц. Ежедневно (например, с 9 до 22 ч) производится измерение нагрузки периодами в 15 мин.

Пусть y_{kj} – интенсивность нагрузки, измеренная в k-й 15-минутный интервал j-го дня. Через λ_{ij} обозначим интенсивность нагрузки в i-й час j-го дня, которая определяется путем усреднения четырех соседних 15-минутных интервалов (табл. 3.1):

$$\lambda_{ij} = \frac{\sum_{k=i}^{i+3} y_{kj}}{4}. \qquad (3.21)$$

Таблица 3.1

День	Интенсивность нагрузки по 15-минутным интервалам				
	9.00–9.15	9.15–9.30	9.30–9.45	9.45–10.00	10.00–10.15	…
j	y_{1j}	y_{2j}	y_{3j}	y_{4j}	y_{5j}	…
	λ_{1j}					
		λ_{2j}				

Полученное распределение интенсивностей нагрузки по часам в дни измерений можно представить в виде табл. 3.2.

Таблица 3.2

День	Интенсивность нагрузки по часовым интервалам			
	9.00–10.00	9.15–10.15	…	21.00–22.00
1	λ_{11}	λ_{21}	…	λ_{n1}
2	λ_{12}	λ_{22}	…	λ_{n2}
…	…	…	…	…
n	λ_{1n}	λ_{2n}	…	λ_{nn}
$\bar{\lambda}_i$	$\bar{\lambda}_1$	$\bar{\lambda}_2$	…	$\bar{\lambda}_n$

Для каждого часа определяют средние за n дней интенсивности нагрузок

$$\bar{\lambda}_i = \frac{\sum_{j=1}^{n} \lambda_{ij}}{n}. \qquad (3.22)$$

Максимальная из величин λ_i и будет интенсивностью в ЧНН:

$$\lambda_{ЧНН} = \max_i \bar{\lambda}_i. \qquad (3.23)$$

Контрольные вопросы:

1. Что называется потоком вызовов?
2. Какие основные характеристики потоков вызовов? Дать определение интенсивности и параметра потока.
3. Каковы принципы классификации потоков вызовов? Дать определение свойств стационарности, ординарности, последействия.
4. Что называется простейшим потоком вызовов? Математическая модель простейшего потока вызовов.
5. Что называется примитивным потоком вызовов? Математическая модель примитивного потока вызовов.
6. Что понимают под поступающей, обслуженной и потерянной нагрузкой?
7. Что называется часом наибольшей нагрузки станции?
8. Назовите основные параметры нагрузки.
9. В каких единицах измеряется нагрузка и интенсивность нагрузки?
10. Приведите формулы для определения нагрузки и поясните значение входящих в них аргументов.

Исходные данные
Для варианта используются ДВЕ цифры номера в журнале

Наименование	Последняя цифра варианта									
	1	2	3	4	5	6	7	8	9	0
Монтируемая ёмкость номеров N	110	800	200	300	600	150	100	110	500	160
Среднее число вызовов на одного абонента в ЧНН: при внутренней связи $C_{проект}$	2,7	2,6	2,1	2,4	2,0	1,8	2,5	2,2	2,3	1,9
по соединительным линиям (сл): к АТС-1 $C_{АТС\ 1}$	0,35	0,45	0,38	0,30	0,55	0,35	0,31	0,5	0,4	0,4
к АТС-2 $C_{АТС\ 2}$	0,55	0,65	0,85	0,6	0,7	0,95	0,75	0,85	0,5	0,9
к АТС-3 $C_{АТС\ 3}$	0,2	0,15	0,21	0,12	0,13	0,2	0,17	0,14	0,18	0,16
к столу справок Cсп	0,04	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,01
к столу заказов Cзак	0,12	0,14	0,14	0,19	0,11	0,17	0,15	0,13	0,18	0,12
Средняя длительность одного разговора при внутренней связи Тпроект,с	60	70	115	110	80	130	100	70	80	90
по с.л. к АТС-1 Т	60	70	60	80	70	100	60	80	70	90
к АТС-2 Т	70	80	110	140	100	120	90	130	100	120
к АТС-3 Т	80	50	60	70	90	120	60	100	90	70
к столу справок Тсп	20	60	25	55	30	50	35	45	40	20
к столу заказов Тзак	22	30	18	27	19	25	20	21	24	15

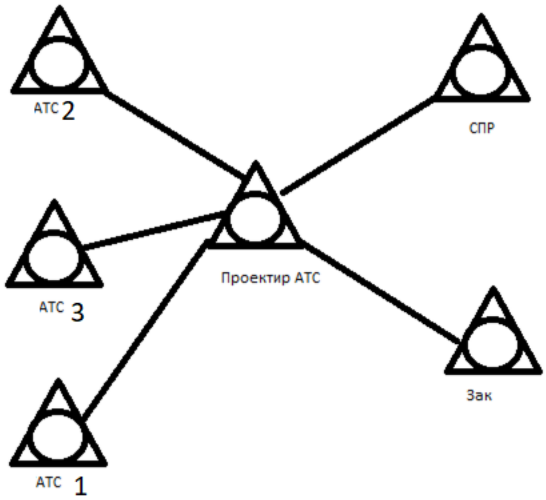
Доля состоявшихся разговоров Кр	0,69	0,6	0,65	0,62	0,64	0,68	0,63	0,66	0,61	0,7
Нагрузка $У_{рсл}$, Эрл, при потерях р	0,002	0,005	0,03	0,05	0,01	0,002	0,005	0,03	0,05	0,01

Наименование процесса	Обозначение	Ср. длительность, с
Слушание сигнала «ответ станции»	$t_{ОС}$	3
Слушание сигнала вызова	$t_{ВЫЗ}$	10
Соединение с момента окончания набора номера до подключении к линии вызываемого абонента	$t_{У}$	2
Набор одного знака номера	$t_{НН}$	1,5

В результате выполнения расчетов необходимо определить нагрузку выполняемую системой коммутации (АТС)

Данное значение необходимо для проектирования сети связи.

На рисунке представлена схема сети связи с проектируемой АТС, а также существующими: городской АТС-1, учережденческой АТС-2, железнодорожной АТС-3 , а также справочная служба и служба заказов.



Параметрами нагрузки являются: число источников нагрузки, среднее число вызовов от одного источника, среднее время занятия приборов. Единица измерения _Эрланг. Эрл.

Нагрузка на DX-500 определяется по формуле:

$$У = \frac{1{,}1 \cdot K_р \cdot N}{3600} \cdot \left(C_{проек'} \cdot Т_{проек} + C_{1'} \cdot Т_1 + C_{2'} \cdot Т_2 + C_{3'} \cdot Т_3\right) + + \frac{N}{3600} \cdot (C_{СПР} \cdot Т_{СПР} + C_{ЗАК} \cdot Т_{ЗАК}). \quad (1)$$

где 1,1 – коэффициент, учитывающий вызовы, не окончившиеся разговором;

С, $К_р$, N – берутся из таблицы исходных данных;

$Т_{1-3}$ – то же, к АТС-1-3;

$Т_{спр}$ – то же, к столу справок;

$Т_{зак}$ – то же, к столу заказов.

Время работы АТС для проектируемой станции:

$$Т_{проект} = Т + t_{уст.соед}\ , \quad (2)$$

$$t_{уст.соед.} = t_{OC} + n * t_{HH} + t_{выз} + t_у \quad (3)$$

где Т – средняя длительность разговора при внутреннем соединении (по таблице исходных данных);

$t_{уст.соед}$ – время установления внутреннего соединения;

n – количество знаков в наборе номера.

Другие значения $T_{(1-3,}$ спр, зак) находятся также как и по предыдущей формуле, только меняется время разговора.

Число знаков в наборе номера городской ГАТС -5, учережденческой УАТС – 3 , железнодорожной ЖАТС - 4, а также справочная служба и служба заказов -2 знака.