

УДК 621.391

МОДЕЛИ И МЕТОДЫ ИССЛЕДОВАНИЯ ПРОЦЕССОВ ФУНКЦИОНИРОВАНИЯ И ОПТИМИЗАЦИИ ПОСТРОЕНИЯ СЕТЕЙ СВЯЗИ СЛЕДУЮЩЕГО ПОКОЛЕНИЯ

А.Н. Назаров, начальник отдела ОАО «Интеллект Телеком», д.т.н., nazarov@i-tc.ru

К.И. Сычев, заместитель начальника отдела НИЦ ФСО России, к.т.н., sychev-k@mail.ru

Ключевые слова: сети связи следующего поколения (NGN), качество обслуживания (QoS), самоподобный трафик.

Введение. Современное развитие информационного общества и информационных систем различного назначения, а также набирающие интенсивность процессы модернизации и технологического развития экономики России предъявляют ряд новых «революционных» требований к телекоммуникационным системам по видам, объемам и качеству передаваемой информации, доступности обслуживания. Основными из них являются:

- интегрированное предоставление услуг и, соответственно, многокомпонентность (мультимедийность) передаваемой информации (голос, видео, данные);
- мобильность пользователей и обслуживания, интерактивность и широкополосность доступа, многосвязное взаимодействие;
- возможность гибкого создания и внедрения новых услуг;
- обеспечение минимальной себестоимости предоставляемых услуг за счет унификации сетевых решений и эффективности использования сетевых ресурсов.

В целом данные факторы определяют эволюционный переход от построения узкоспециализированных выделенных (по видам связи) сетей к мультисервисным сетям связи с пакетной IP, MPLS (ATM) коммутацией, реализуемым на основе концепции NGN. Последние обеспечивают предоставление неограниченного набора услуг с гибкими возможностями по их управлению, персонализации и созданию новых услуг за счет унификации сетевых решений на основе разделения функций переноса, коммутации, управления вызовами и услугами. Это достигается [1, 2]:

- четырехуровневой архитектурой построения сетей NGN, включающей уровень доступа, транспортный уровень, уровень управления вызовами, уровень услуг, а также применением на каждом уровне открытых стандартов;
- обеспечением QoS при передаче разнородной информации на основе введения классов качества.

Изложенные особенности развития инфокоммуникационных услуг и телекоммуникационных систем требуют адекватного развития моделей и методов исследования процессов функционирования и оптимизации построения (проектирования) сетей связи. Разработку данных моделей и методов для сетей NGN будем вести в области приоритетных систем обслуживания с произвольными распределениями поступления и обслуживания заявок различных классов качества, ограничениями ресурсов (объем буфера, число каналов и их пропускная способность). Это позволит при решении поставленных задач учитывать специфику совместной передачи разнородной информации с различными требованиями к QoS, а также пачечную (самоподобную) природу трафика сетей NGN.

С учетом реальных применений сетевых протоколов будем считать, что приоритеты в обслуживании являются отно-

сительными (без прерывания обслуживания), назначаются по классам качества (видам) услуг, категориям пользователей и подсистемам (информационного обмена, сигнализации, управления и др.) и определяются задержками и потерями пакетов. Общее число приоритетов $r = 1, R$.

Разработку данных моделей начнем с классического случая узла коммутации (УК) типа $\bar{M}_r / M / V / K$ с простейшими потоками, что обусловлено фундаментальностью получаемых результатов и возможностью распространения последних на более общие случаи построения моделей систем обслуживания.

Модель узла коммутации типа $\bar{M}_r / M / V / K$. Пусть исследуемый УК состоит из $1 \leq V < \infty$ одинаковых обслуживающих приборов с пропускной способностью, определяемой номинальным рядом скоростей $\{V \times 2048 \text{ кбит/с}, \dots\}$, и буфера объемом $0 < K < \infty$. На вход рассматриваемого УК поступают $r = 1, R$ простейших потоков пакетов с интенсивностью λ_r , обслуживание которых осуществляется с относительными приоритетами. Длительность обслуживания пакетов приоритетов $r = 1, R$ определяется показательным распределением с параметром μ : $B(t) = 1 - \exp(-\mu t)$, $\mu \geq 0, t \geq 0$.

В [3] для модели УК данного типа получено стационарное распределение вероятностей состояний i ($0 \leq i \leq V + K$):

$$P_i(r, V) = \begin{cases} \frac{\rho_r^i}{i!} P_0(R, V), & 0 \leq i < V; \\ Q \left(\frac{\rho_r}{V} \right)^{i-V} \left[1 - \frac{\rho_r}{V} \right] \left[1 - \left(\frac{\rho_r}{V} \right)^{K+1} \right]^{-1}, & \rho_r \neq V, V \leq i \leq V + K; \\ \frac{Q}{K+1}, & \rho_r = V, V \leq i \leq V + K. \end{cases} \quad (1)$$

Это выражение позволяет определить необходимые показатели качества функционирования:

- вероятность потери пакетов r -го приоритета – состояние УК: $i = V + K$;
- среднюю длину очереди пакетов r -го приоритета (рекуррентным способом):

$$l_r = \begin{cases} L_r, & r = 1; \\ L_r - L_{r-1}, & r = 2, R; \end{cases} \quad L_r = \sum_{m=1}^K m P_{V+m}(r, V); \quad (2)$$

- среднее время ожидания обслуживания пакетами r -го приоритета:

$$w_r = l_r / \lambda_r; \quad (3)$$

- среднее время пребывания пакетов r -го приоритета в УК, рассчитываемое с учетом независимости времени ожидания w_r и обслуживания $b = \mu^{-1}$ пакетов:

$$t_{\text{ин}}(r) = w_r + b. \quad (4)$$

В выражении (1) $P_0(R, V)$ – вероятность того, что в УК на обслуживании отсутствуют пакеты приоритетов $r = 1, R$, определяется из условий нормировки:

$$P_0(R, V) = \left[\sum_{m=0}^V \frac{\rho_R^m}{m!} + \frac{\rho_R^V}{V!} \sum_{m=1}^K \left(\frac{\rho_R}{V} \right)^m \right]^{-1};$$

$Q = 1 - \sum_{i=0}^{V-1} P_i(R, V)$ — вероятность того, что все каналы УК заняты обслуживанием суммарного потока пакетов приоритетов $r = \overline{1, R}$; $\rho_r = \frac{\Lambda_r}{\mu}$, $\Lambda_r = \sum_{i=1}^r \lambda_i$ — интенсивность суммарной нагрузки пакетов с приоритетами $\overline{1, r}$; L_r — средняя суммарная длина очереди пакетов приоритетов $\overline{1, r}$. При этом ввиду конечности числа состояний i ($0 \leq i \leq V + K$) рассматриваемого УК условие стационарности процессов его функционирования определяется для любых значений ρ_r ($0 < \rho_r < \infty$).

В [2] для УК типа $\overline{M}_r / M / V / K$ получено стационарное распределение времени ожидания пакетов различных приоритетов (в форме преобразования Лапласа-Стилтjesа), характеризующее распределением Эрланга:

$$w_r(s) = \sum_{i=0}^{K-1} P_{V+i}(r, V) \left(\frac{\mu V}{s + \mu V} \right)^{i+1} = \mu V P_V(r, V) \frac{1 - \left(\frac{\Lambda_r}{s + \mu V} \right)^K}{s + \mu V - \Lambda_r}, \quad (5)$$

что также позволяет рекуррентным способом определить среднее время ожидания пакетов по приоритетам:

$$w_r = \begin{cases} w_1^*, & r = 1; \\ \frac{\Lambda_r}{\lambda_r} \left(w_r^* - \frac{\Lambda_{r-1}}{\lambda_r} w_{r-1}^* \right), & r = \overline{2, R}. \end{cases} \quad (6)$$

В выражении (6) w_r^* — среднее время ожидания обслуживания в бесприоритетном УК с суммарной входящей нагрузкой ρ_r (без учета приоритетов):

$$w_r^* = -w_r'(s=0) = \begin{cases} P_V(r, V) \frac{1 - (K+1) \left(\frac{\rho_r}{V} \right)^K + K \left(\frac{\rho_r}{V} \right)^{K+1}}{\mu V \left(1 - \frac{\rho_r}{V} \right)^2}, & \frac{\rho_r}{V} \neq 1; \\ P_V(r, V) \frac{K(K+1)}{2\mu V}, & \frac{\rho_r}{V} = 1. \end{cases} \quad (7)$$

Исходя из функции распределения времени ожидания пакетов (5), можно дополнительно определить следующие показатели качества функционирования УК:

- джиттер задержки пакетов (для трафика реального времени)

$$\sigma(t_{\text{ин}}(r)) = \sqrt{t_{\text{ин}}^{(2)}(r) - (t_{\text{ин}}(r))^2} \quad (8)$$

- вероятность того, что пакет r -го приоритета, поступивший в произвольный момент времени в УК типа $\overline{M}_r / M / V / K$, попадает на ожидание и среднее время его ожидания не превысит допустимого значения:

$$P(w_r \leq t_{\text{ин}}(r)) = 1 - \frac{P_V(r, V)}{\exp(\mu V t_{\text{ин}}(r))} \sum_{i=0}^{K-1} \left(\frac{\rho_r}{V} \right)^i \sum_{j=0}^i \frac{(\mu V t_{\text{ин}}(r))^j}{j!}; \quad (9)$$

- вероятность своевременной доставки (ВСД) пакетов r -го приоритета

$$P[t_{\text{ин}}(r) \leq t_{\text{ин}}(r)] = P(w_r \leq t_{\text{ин}}(r)) P(b \leq t_{\text{ин}}(r)), \quad (10)$$

- являющуюся комплексным показателем, характеризующим совместное выполнение требований по задержкам пакетов при передаче разнородного трафика.

В выражениях (9)–(10): $t_{\text{ин}}(r)$ — требования к среднему времени доставки пакетов r -го приоритета, нормируемые рекомендациями МСЭ-Т Y.1540 и др.;

$$P(b \leq t_{\text{ин}}(r)) = \int_0^{t_{\text{ин}}(r)} dB(t)$$

— вероятность того, что среднее время обслуживания пакетов не превысит допустимого значения;

$$t_{\text{ин}}^{(2)}(r) = w_r^{(2)} + 2w_r b + b^{(2)} \quad (11)$$

— второй момент среднего времени пребывания пакетов r -го приоритета в УК [2, 4]; $w_r^{(2)}$ — второй момент времени ожидания пакетов r -го приоритета в УК, определяемый распределением (5) по аналогии с рекуррентным выражением (6).

Модель узла коммутации NGN типа $\overline{G}_r / G / V / K$. Для исследования процессов функционирования УК сетей NGN типа $\overline{G}_r / G / V / K$ при произвольных распределениях поступления и обслуживания заявок различных классов качества (приоритетов) [2] воспользуемся результатами (1)–(11), а также:

- методом диффузионной аппроксимации [5, 6], обеспечивающим в условиях большой нагрузки получение решения для УК общего типа $G / G / V / \infty$ — при произвольных распределениях поступления и обслуживания заявок (без приоритетов), исходя из решения для УК «классического» типа $M / M / V / K$, $K \leq \infty$;

- свойством эквивалентности систем обслуживания с конечным и бесконечным буфером при малых нормах потерь пакетов [5], обеспечивающим для расчета требуемых показателей качества возможность применения моделей обслуживания, как с конечным, так и бесконечным буфером;

- «законом сохранения накопленной в очереди работы» [4, 6], который утверждает, что для приоритетных дисциплин, не допускающих прерывания обслуживания и простоя обслуживаемых приборов, накопленная в очереди работа постоянна и равна накопленной работе в бесприоритетной системе обслуживания с суммарной нагрузкой. Данный «закон» распространяется на системы обслуживания с относительными приоритетами, что позволяет рекуррентно, по аналогии (2), (6), рассчитывать требуемые показатели качества функционирования.

Определим необходимые показатели качества функционирования УК сетей NGN при передаче разнородной информации.

1. На основании метода диффузионной аппроксимации в [5, 6] получено выражение для средней длины очереди в УК общего типа $G / G / V / \infty$:

$$L = P_{i>V-1} \frac{\rho}{1-\rho} \frac{C_a^2 + C_b^2}{2}; \quad \rho \equiv \rho_R, \quad \frac{\rho}{V} < 1, \quad (12)$$

где $P_{i>V-1}$ характеризует вероятность того, что пакет, поступающий в систему, застанет все обслуживающие приборы занятыми (определяется второй формулой Эрланга), а C_a^2, C_b^2 — соответственно, квадратичные коэффициенты вариации для распределений входящего потока и времени обслуживания пакетов.

При этом решение (12) не зависит от законов распределения входящего потока и времени обслуживания пакетов и определяется только первыми двумя моментами данных распределений. При $V=1, C_a^2 = 1$ выражение (12) сходится к формуле Поллячека-Хинчина [6]. С учетом этого выражение

(12) можно переписать в виде произведения средней длины очереди L_k в «классической» системе обслуживания типа $M/M/V/\infty$ на коэффициент $(C_a^2 + C_b^2)/2$, характеризующий среднюю изменчивость распределений входящего потока и времени обслуживания пакетов:

$$L = L_k \frac{C_a^2 + C_b^2}{2}; \quad L_k = P_{i>V-1} \frac{\rho}{1-\rho}. \quad (13)$$

Для упрощения и сокращения представления последующих выражений введем обозначение $C_{ab}^2 = (C_a^2 + C_b^2)/2$. С целью определения средней длины очереди пакетов по приоритетам предварительно перепишем выражение (12) на основании формулы Литтла:

$$L = \Lambda w C_{ab}^2; \quad w = \frac{P_{i>V-1}}{\mu(1-\rho)}, \quad (14)$$

где w — среднее время ожидания обслуживания в УК типа $M/M/V/\infty$ с суммарной входящей нагрузкой $\rho \equiv \rho_R$ (без учета приоритетов).

Тогда, согласно «закону сохранения накопленной в очереди работы» и свойству эквивалентности УК с конечным и бесконечным буфером при малых нормах потерь пакетов, средняя суммарная длина очереди пакетов с приоритетами l, r определяется выражением:

$$L_r = \Lambda_r w_r^* C_{ab}^2, \quad (15)$$

где w_r^* — среднее время ожидания пакетов в УК с суммарной входящей нагрузкой ρ_r (без учета приоритетов), определяемое выражением (7).

Полученное выражение (15) позволяет также на основании «закона сохранения накопленной в очереди работы» рекуррентным способом рассчитать среднюю длину очереди пакетов r -го приоритета в УК типа $\bar{G}_r/G/V/K$:

$$l_r = \begin{cases} L_r, & r = 1; \\ L_r - L_{r-1}, & r = \overline{2, R}. \end{cases} \quad (16)$$

2. Среднее время ожидания пакетов r -го приоритета в УК типа $\bar{G}_r/G/V/K >$ определяется из (16) на основании формулы Литтла:

$$w_r = l_r / \lambda_r. \quad (17)$$

3. Джиттер задержки пакетов $\sigma(t_{nn}^{(r)})$ в УК типа $\bar{G}_r/G/V/K$ определяется зависимостью (8), в которой второй момент среднего времени пребывания пакетов r -го приоритета $t_{nn}^{(2)}(r)$ можно вычислить по аналогичной (11) формуле:

$$t_{nn}^{(2)}(r) = (C_{ab}^2)^2 w_r^{(2)} + 2w_r b + b^{(2)}, \quad (18)$$

где $w_r^{(2)}$ определяется распределением (5) по аналогии с рекуррентным выражением (6); w_r — выражением (17); b и $b^{(2)}$ — распределением времени обслуживания пакетов.

4. На основании метода диффузионной аппроксимации в [5] представлено выражение для вероятности потери пакетов в УК общего типа $G/G/1/K$ (состояние УК: $i=K+1$):

$$P_{K+1} = \frac{1-\rho}{1-\rho^{K/C_{ab}^2+1}} \rho^{K/C_{ab}^2+1}, \quad \rho \equiv \rho_R, \quad \rho < 1. \quad (19)$$

Это выражение также как и (12) не зависит от законов распределения входящего потока и времени обслуживания пакетов (определяется только первыми двумя моментами данных распределений) и при $C_a^2 = C_b^2 = 1$ сходится к формуле (1) для «классической» системы обслуживания типа $M/M/1/K$.

Для определения вероятности потери пакетов различных

приоритетов в УК типа $\bar{G}_r/G/V/K$ предварительно перепишем выражение (18) в форме:

$$P_{K+1} = P_{i>0} \frac{1-\rho}{1-\rho^{K/C_{ab}^2+1}} \rho^{K/C_{ab}^2}. \quad (20)$$

Выражение (20) характеризует вероятность выполнения совместного события: обслуживающий прибор (с вероятностью $P_{i>0} = \rho$) и буфер заняты обслуживанием пакетов.

Перепишав с учетом (1) выражение (20) для случая $V > 1$:

$$P_{V+K} = P_{i>V-1} \frac{1-\rho/V}{1-(\rho/V)^{K/C_{ab}^2+1}} (\rho/V)^{K/C_{ab}^2}, \quad \rho/V < 1, \quad (21)$$

можно определить искомую вероятность потери пакетов в УК (по приоритетам):

$$P_{V+K}(r) = P_{i>V-1}(R, V) \frac{1-\rho_r/V}{1-(\rho_r/V)^{K/C_{ab}^2+1}} (\rho_r/V)^{K/C_{ab}^2}. \quad (22)$$

Последнее выражение, как и в случае (20), характеризует выполнение совместного события, заключающегося в том, что:

- обслуживающие приборы заняты обслуживанием суммарного потока пакетов l, R , вероятность данного события $P_{i>V}(R, V) = Q$ определяется распределением (1), полученным для экспоненциального времени обслуживания пакетов;
- буфер занят пакетами с более высоким и/или равным приоритетом.

5. Вероятность своевременной доставки пакетов r -го приоритета в УК типа $\bar{G}_r/G/V/K$ определяется на основе (10) выражением:

$$P[t_{nn}^{(r)}(r) \leq t_{nn\text{тр}}^{(r)}(r)] = \int_0^{t_{nn\text{тр}}^{(r)}} dW_r(t) \int_0^{t_{nn\text{тр}}^{(r)}} dB(t). \quad (23)$$

Определим неизвестную функцию распределения $W_r(t)$ времени ожидания пакетов r -го приоритета в УК типа $\bar{G}_r/G/V/K$. В общем случае, согласно теории диффузионной аппроксимации (для условий большой загрузки) [6], распределение времени ожидания пакетов $W(t)$ в УК общего типа $G/G/V/\infty$ (без приоритетов) имеет показательный характер и с учетом принятого приближения (14) может быть определено как:

$$W(t) \approx 1 - \exp\left(-\frac{\mu(1-\rho)}{P_{i>V-1}} C_{ab}^2 t\right), \quad t \geq 0. \quad (24)$$

Тогда, согласно «закону сохранения» и свойству эквивалентности УК с конечным и бесконечным буфером при малых нормах потерь пакетов, перепишем выражение (24) с целью определения функции распределения времени ожидания пакетов $W_r^*(t)$ в УК типа $\bar{G}_r/G/V/K$ с суммарным трафиком приоритетов l, r :

$$W_r^*(t) \approx 1 - \exp\left(-\frac{C_{ab}^2}{w_r^*} t\right), \quad t \geq 0, \quad (25)$$

где w_r^* — среднее время ожидания обслуживания в УК с суммарной входящей нагрузкой ρ_r (без учета приоритетов), определяемое выражением (7).

Исходя из (25), функция распределения времени ожидания пакетов r -го приоритета $W_r(t)$ в УК типа $\bar{G}_r/G/V/K$ определяется на основе рекуррентного выражения:

$$W_r(t) = \begin{cases} W_1^*(t), & r = 1, \\ \frac{\Lambda_r}{\lambda_r} \left(W_r^*(t) - \frac{\Lambda_{r-1}}{\Lambda_r} W_{r-1}^*(t) \right), & r = \overline{2, R}. \end{cases} \quad (26)$$

Функцию распределения времени обслуживания пакето-

тов $B(t)$ будем определять из необходимости расчета верхних и нижних граничных оценок показателей качества функционирования УК (на наихудший и наилучший случаи), задаваемых, соответственно, экспоненциальным ($C_b^2 = 1$ — переменная длина пакета) и детерминированным ($C_b^2 = 0$ — фиксированная длина пакета) временем обслуживания пакетов. При этом предлагаемые верхние и нижние граничные оценки показателей качества функционирования гарантированно включают основные варианты распределений длительности обслуживания пакетов для УК сетей NGN на основе технологий ATM, IP, MPLS. Последние — наихудший случай — (IP, MPLS) представляют собой ступенчатую функцию, характеризующую вероятностной смесью вырожденных распределений — каждое с собственной фиксированной длительностью пакета. Данная ступенчатая функция может быть с заданной точностью аппроксимирована смесью Эрланговских распределений [7], которая, в свою очередь, имеет квадратичный коэффициент вариации $C_b^2 = [0, 1]$, удовлетворяющий предложенным границам.

Обслуживание самоподобного трафика. В настоящее время многочисленными исследованиями доказано, что реальный сетевой трафик является самоподобным (фрактальным) случайным процессом [8-10]. При этом на самоподобность сетевого трафика оказывают влияние следующие факторы: изменчивость поведения пользователей, неравномерная структура передаваемых данных, объединение трафика и ограниченность сетевых ресурсов, механизмы сетевого управления, развитие услуг (переменная скорость передачи, мультимедиа), а также их различные корреляции. В результате статистические характеристики самоподобного сетевого трафика проявляют свойства масштабирования, долговременной зависимости и медленно убывающей дисперсии, которые сохраняются при различных операциях в сети, таких как статистическое мультиплексирование, буферизация. Следствием данных свойств являются степенные законы изменения статистических характеристик очереди и задержек, а также необходимость описания этих процессов с помощью распределений с «тяжелыми хвостами».

В [10, 11] показано, что для исследования асимптотических самоподобных процессов в большинстве практических случаев могут быть использованы модели ON/OFF-источников. Длительность интервалов ON/OFF-периодов имеет распределение Парето (вопросы исследования процессов обслуживания с учетом классов QoS строго самоподобного трафика, моделируемого на основе фрактального броуновского движения, рассмотрены в [2]). Вместе с тем распределение Парето с характеристическим показателем $1 < \alpha < 2$ обладает бесконечной дисперсией, что обусловлено отсутствием ограничений на максимальный размер передаваемых блоков данных и делает невозможным получение устойчивых оценок для показателей качества функционирования УК.

Однако в реальных сетях связи размеры передаваемых протокольных блоков данных (пачек пакетов) ограничены. Кроме того, факторами, ограничивающими размеры пачек пакетов, являются конечные размеры буферов УК, сетевые политики обеспечения QoS.

В соответствии с этим в [11] для распределения Парето $\omega(x)$ введены пределы интегрирования $\int_{\chi_{\min}}^{\chi_{\max}} \omega(x) dx = 1$, позволяющие получить плотность распределения вероятностей для ограниченного (усеченного) распределения Парето:

$$\omega_o(x) = \frac{\alpha \chi_{\min}^{\alpha} x^{\alpha-1}}{1 - \left(\frac{\chi_{\min}}{\chi_{\max}}\right)^{\alpha}} \quad (27)$$

$$\chi_{\min} \leq x \leq \chi_{\max}, \quad 0 < \alpha < 2, \quad \alpha = 3 - 2H,$$

где χ_{\min}, χ_{\max} — минимальное и вводимое максимальное значение случайной величины, характеризующее размеры протокольных блоков данных (пачек пакетов) или интервалы времени между их приходами; H — показатель Херста.

Распределение $\omega_o(x)$, как и исходное распределение $\omega(x)$, является распределением с «тяжелым хвостом», имеющим квадратичный коэффициент вариации $C_a^2 > 1$. При этом проведенные с использованием имитационного моделирования исследования [2] показали, что наименьшей относительной погрешностью для оценок средней длины очереди при обслуживании потоков, описываемых ограниченным распределением Парето, обладает приближение (12). Это позволяет, с учетом полученных результатов (15)–(18), (22)–(26), однозначно перейти к исследованию процессов функционирования аналогичных УК с асимптотически самоподобным трафиком путем соответствующего выбора квадратичного коэффициента вариации входящего потока:

$$C_a^2 = \left(\frac{\chi_{\max}^2 \chi_{\min}^{\alpha_r} - \chi_{\max}^{\alpha_r} \chi_{\min}^2}{2 - \alpha_r} - \frac{\alpha_r (\chi_{\max}^{\alpha_r} \chi_{\min} - \chi_{\max} \chi_{\min}^{\alpha_r})^2}{(1 - \alpha_r)^2 (\chi_{\max}^{\alpha_r} - \chi_{\min}^{\alpha_r})} \right) \times \times \frac{(1 - \alpha_r)^2 (\chi_{\max}^{\alpha_r} - \chi_{\min}^{\alpha_r})}{\alpha_r (\chi_{\max}^{\alpha_r} \chi_{\min} - \chi_{\max} \chi_{\min}^{\alpha_r})^2}, \quad \alpha_r = 3 - 2H_r, \quad (28)$$

где χ_{\min}, χ_{\max} — минимальное и вводимое максимальное значение случайной величины, характеризующее размеры пачек пакетов для объединенного трафика r -го приоритета (класса обслуживания); α_r, H_r — соответственно, характеристический показатель и показатель Херста для объединенного трафика r -го приоритета (класса обслуживания).

Обслуживание произвольного трафика. Для получения непосредственного решения относительно произвольных распределений времени поступления и обслуживания пакетов в УК типа $\bar{G}_r / G / V / K$ необходимо собрать и обработать статистические данные по распределениям интервалов времени между поступлениями пакетов и длин пакетов b_r (по приоритетам) и на основе моментов данных распределений рассчитать искомые значения квадратичных коэффициентов вариации C_a^2, C_b^2 для подстановки в формулы (15)–(18), (22)–(26).

Назначение приоритетов. Выше показано, что в сетях NGN для обеспечения совместной передачи сообщений различных классов с требуемым для каждого из них качеством вводятся относительные приоритеты. Оптимизация назначения приоритетов позволяет минимизировать потери (затраты), связанные с задержками пакетов различных классов в УК [4]. В соответствии с этим в [2] доказано, что оптимальный порядок обслуживания определяется исходя из первого момента длительности обслуживания пакетов различных классов b_r , а также требований к времени их доставки в сети, обоснованных нормативными документами МСЭ-Т:

$$\frac{c_1^*}{b_1^*} \geq \frac{c_2^*}{b_2^*} \geq \dots \geq \frac{c_r^*}{b_r^*} \geq \frac{c_R^*}{b_R^*}, \quad c_r^* = f(t_{\text{нп тр}}(r)), \quad b_r^* = f(b_r), \quad (29)$$

где $f(t_{\text{нп тр}}(r)), f(b_r)$ — функции нормализации значений $t_{\text{нп тр}}(r), b_r$.

Неравенство (29) характеризует присвоение приоритета коротким пакетам с жесткими требованиями к времени доставки в сети. При этом, согласно (29), оптимальное на-

значение приоритетов не зависит от интенсивностей потоков пакетов и их статистической структуры, а также от вида распределения времени обслуживания пакетов [4]. Данное решение в [2] получено для общего случая — когда функция распределения времени обслуживания пакетов $Bv(t)$ и, соответственно, среднее время обслуживания пакетов $b\bar{v}$ различны для различных классов качества обслуживания (приоритетов).

Исследование свойств разработанных моделей УК. Для разработанных моделей УК (P - при $R=2, \rho_1 = \rho_2, V=1, K=50$) построены графические зависимости основных показателей качества функционирования при обслуживании простейших и асимптотически самоподобных потоков (обозначение входящего потока на основе ограниченного распределения Парето):

- средняя длина очереди пакетов (рис. 1, 2);
- вероятность потери пакетов (рис. 1, 3) и др. [2].

Эти характеристики позволяют определять верхние и нижние граничные оценки показателей качества функционирования УК NGN (на наихудший и наилучший случаи), задаваемые, соответственно, экспоненциальным $C_b^2 = 1$ и детерминированным $C_b^2 = 0$ временем обслуживания пакетов.

При этом характер полученных зависимостей показывается:

- в случае самоподобного трафика средняя длина очереди и потери пакетов различных классов обслуживания значительно возрастают по сравнению с простейшим входящим потоком (рис. 1);
- управление QoS на основе относительных приоритетов не устраняет резкого возрастания очереди и потерь пакетов при большой пачечности трафика (рис. 2, 3).

- рис. 1. - средняя длина очереди пакетов и вероятность потери пакетов (при $K \rightarrow \infty$) в УК сетей NGN (без приоритетов), определяемые выражениями (12), (19)

Сетевые модели. Такие модели обеспечивают оценку сквозного QoS в сетях NGN. При этом сети NGN предлагается моделировать на основе приближенного метода декомпозиционной аппроксимации с помощью модели открытой смешанной сети связи с r -классами сообщений, узлы которой моделируются системами обслуживания типа $\bar{G}_r / G / V / K$. В общем случае при использовании метода декомпозиционной аппроксимации моделирование процессов функционирования сетей NGN включает следующие этапы [2]:

- задание графа сети $G(N, M)$, где N, M — множество вершин и ребер;
- решение задач маршрутизации и распределения потоков (по приоритетам);
- декомпозиция сети на независимые элементы — УК, соответствующие линиям связи $j = 1, M$;
- расчет показателей качества функционирования на заданных направлениях связи $i = 1, I$, включая:

а) среднее время доставки сообщений различных приоритетов:

$$t_{дс_i}(r) = t_{уст\ c_i} + \left[t_o(r) + \sum_{q(r)} \sum_{k \in i} (t_{пн_k}(r) + t_{п_k}) \right], \quad r = \bar{1}, \bar{R}, \quad (30)$$

где $t_{уст\ c_i}$ — среднее время установления (разъединения) соединения в i -м направлении связи (для постоянных виртуальных каналов $t_{уст\ c_i} = 0$); $t_o(r)$ — среднее время обработки (кодирования и декодирования, сегментации и сборки — выравнивающая задержка) пакетов; $q(r)$ — среднее число паке-

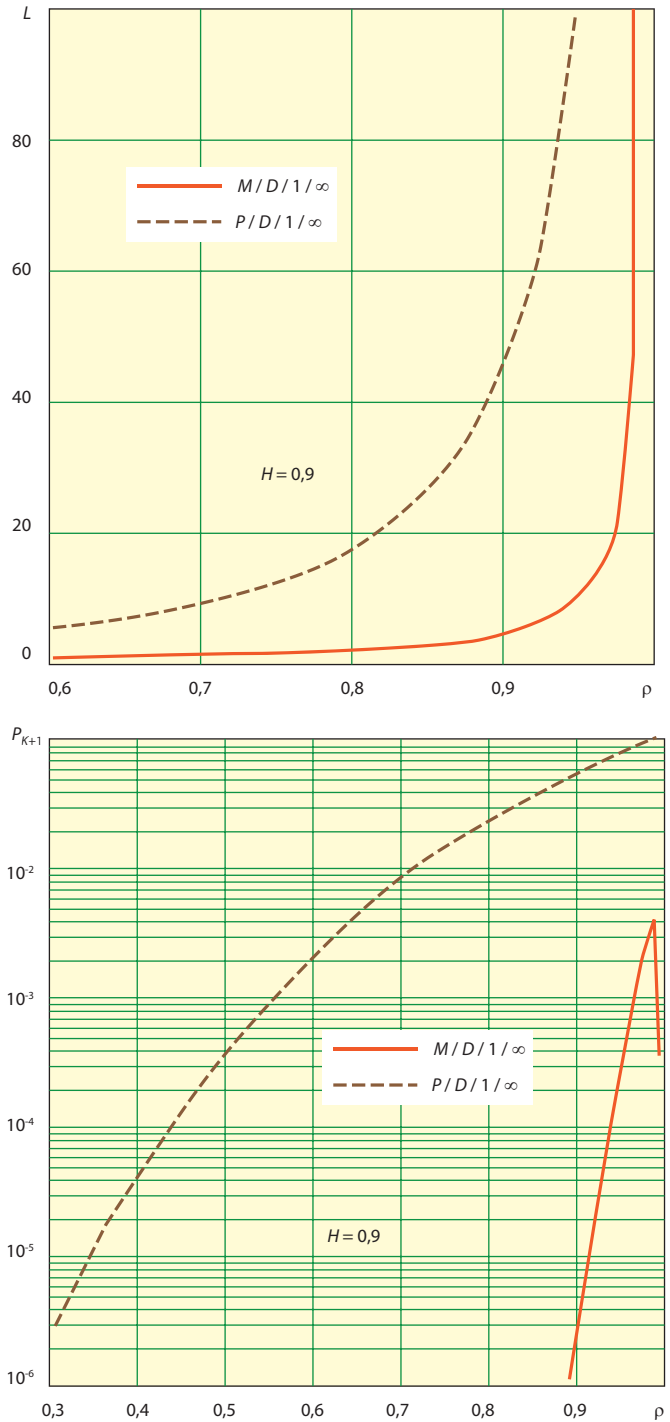


Рис. 1

тов в сообщении $г$ -го приоритета; $t_{п_k}$ — задержка распространения; k — число интервалов связи в i -м направлении связи; б) суммарная вероятность потери пакетов различных приоритетов при передаче в сети (сетевые потери пакетов):

$$P_{нет_i}(r) = \sum_{k \in i} P_{V+K,k}(r); \quad (31)$$

в) вероятность своевременной доставки сообщений различных приоритетов:

$$P[t_{дс_i}(r) \leq t_{дс\ тр}(r)] = P[t_o(r) \leq t_{о\ доп}(r)] \prod_{q(r)} \prod_{k \in i} P[t_{пн_k}(r) \leq t_{пн\ тр}(r)], \quad (32)$$

где $t_{о\ доп}(r)$ — допустимое время обработки пакетов $г$ -го при-

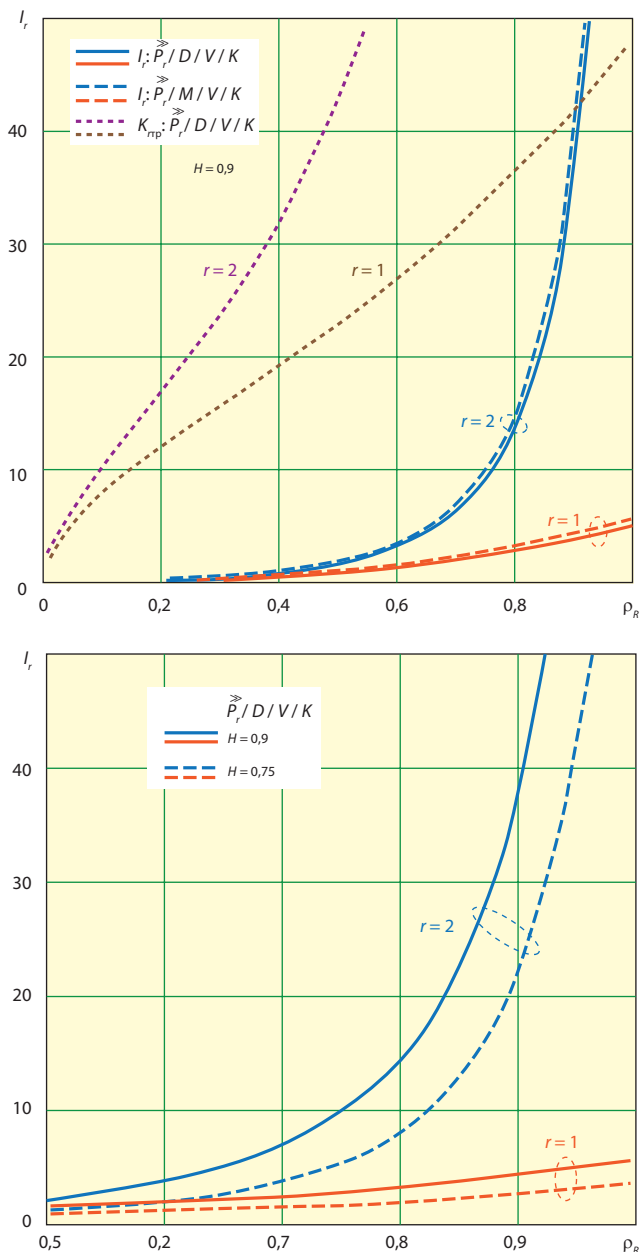


Рис. 2

оритета в оконечных устройствах на передающей и приемной стороне; $t_{дс}(r)$ — среднее время доставки сообщений r -го приоритета; $t_{дс\text{ тр}}(r)$ — требуемое (нормативное) время доставки сообщений r -го приоритета (время старения сообщений r -го приоритета), характеризующее эффективность (возможность) использования доставляемых сетью NGN сообщений по назначению.

При этом среднесетевая вероятность своевременной доставки сообщений является обобщенным показателем эффективности функционирования сети NGN, используемым при оптимизации пропускной способности каналов передачи:

$$P_{\Phi} = \frac{1}{\Lambda^*} \sum_{i=1}^I \sum_{r=1}^R \lambda_i(r) P(t_{дс_i}(r) \leq t_{дс\text{ тр}}(r)); \quad (33)$$

$$\Lambda^* = \sum_{i=1}^I \sum_{r=1}^R \lambda_i(r).$$

Метод оптимизации пропускной способности каналов передачи сетей NGN. Для оптимизации пропускной способности

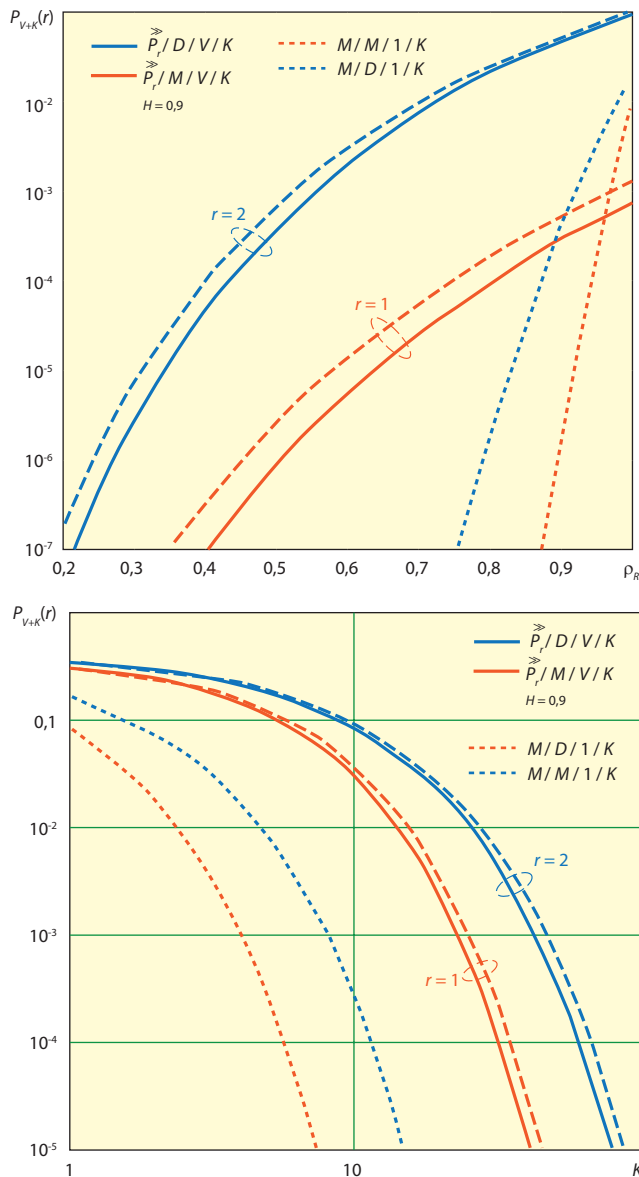


Рис. 3

каналов передачи

$$\mathbf{V}^{(0)} = [V_1^{(0)}, \dots, V_M^{(0)}]^T,$$

$j = \overline{1, M}$ будем решать задачу минимизации затрат $\mathbf{C} = \delta \mathbf{V} + \mathbf{K}_3$, включающих стоимость аренды каналов передачи $\delta \mathbf{V}$ и приведенную стоимость вновь устанавливаемого оборудования сети \mathbf{K}_3 (капитальные затраты), при совместном выполнении требований к: ВСД пакетов $P_{дп\text{ тр}}(r)$, сообщений $P_{дс\text{ тр}}(r)$, $P_{дс\text{ тр}}$; джиттеру задержки пакетов $\sigma_{\text{тр}}(r)$ для трафика реального времени; потерям пакетов $P_{нет\text{ тр}}(r)$ различных приоритетов:

$$\min_{\mathbf{V} \in \Omega_{\chi}} \mathbf{C} \Rightarrow \mathbf{V}^{(0)};$$

$$P[t_{\text{ин}j}(r) \leq t_{\text{ин}\text{ тр}}(r)] \geq P_{дп\text{ тр}}(r) \vee P_{нетj}(r) \leq P_{нет\text{ тр}}(r); \quad (34)$$

$$P_{\Phi_j}(r) \geq P_{дс\text{ тр}}(r), \quad j = \overline{1, M}, \quad P_{\Phi} \geq P_{дс\text{ тр}},$$

$$\sigma_j(t_{\text{ин}}(r)) \leq \sigma_{\text{тр}}(r).$$

Так как зависимости затрат и ВСД пакетов (сообщений) от числа каналов представляют собой монотонные ступенчатые функции, то для решения задачи (34) предлагается итерационный метод прямого поиска, основанный на расчете

чувствительности целевой функции и управлению по состояниям.

В основу метода положена следующая идея: поочередное изменение числа каналов в линиях связи приводит к неодинаковому изменению целевой функции и функций ограничений. Поэтому на очередном шаге поиска меняется лишь тот компонент вектора оптимизируемых переменных, который обеспечивает наибольшее изменение целевой функции. В соответствии с этим предлагаемый метод включает следующие шаги.

1. Выбор начальной точки оптимизации $\mathbf{V}^{(h)}$ на основе метода инвариантов — «закона сохранения обслуженной нагрузки»:

$$V_j^{(h)} \approx \left| \sum_r \rho_{r_j}^* P_{\text{дп тр}}(r) \right| \leq V_j^{(0)}, \quad 0 < \rho_{r_j}^* = \frac{\lambda_{j,r}}{\mu} < \infty. \quad (35)$$

2. На k -ом шаге расчет для каждой $j = \overline{1, M}$ линии связи для значений числа каналов $V_j^{(k)}$ и $V_j^{(k)} + 1$ требуемого объема буфера (определяется решением уравнения (21) относительно K), при котором потери пакетов различных приоритетов будут соответствовать нормам $P_{\text{нет тр}}(r)$, заданным в рекомендациях МСЭ-Т и в соглашении по трафику (см. рис. 2):

$$K_{r_p} = \left[\frac{C_{ab}^2}{\ln(\rho_r / V)} \ln \left(\frac{P_{\text{нет тр}}(r)V}{P_{\text{нет тр}}(r)\rho_r + P_{i>V}(R,V)V - P_{i>V}(R,V)\rho_r} \right) \right]. \quad (36)$$

На основе (36) рассчитывают ВСД пакетов (23), сообщений (32), (33), джиттера задержки пакетов (8) и относительного приращения целевой функции на единицу затрат:

$$\gamma_j(V_j^{(k)}) = \frac{P_{\Phi_j}(V_j^{(k)} + 1) - P_{\Phi_j}(V_j^{(k)})}{C_j P_{\Phi_j}(V_j^{(k)})}. \quad (37)$$

3. Увеличение пропускной способности той линии связи, для которой требования по ВСД не выполнены и обеспечивается максимальное относительное приращение ВСД на единицу затрат: $\gamma_j(V_j^{(k)}) = \inf_{1 \leq m \leq M} \gamma_m(V_m^{(k)})$.

4. Определение текущих значений (с учетом изменений на шаге 3) показателей качества $P[t_{\text{пт}j}(r) \leq t_{\text{пт тр}}(r)]$, $\sigma_j(t_{\text{пт}}(r))$, $P_{\Phi_j}(r)$, P_{Φ} и затрат C .

5. Проверка выполнения условий задачи (требований к показателям качества) и достаточности ресурсов (ограничений по стоимости $C \leq C_{\text{max}}$, пропускной способности $\mathbf{V}^{(k+1)} \in \Omega_{\chi}$). При их выполнении оптимизация завершается, в противном случае переход к шагу 2.

Заключение. В статье представлены новые модели и методы исследования процессов функционирования и оптимизации построения сетей NGN.

В целом их можно использовать для проектирования (оп-

тимизации построения) сетей NGN на основе модели как дифференцированных услуг (при назначении приоритетов в обслуживании для подсистем информационного обмена, сигнализации, управления), так и интегрированных услуг (при последовательном проектировании виртуальных наложенных сетей информационного обмена (с приоритетами по видам услуг и категориям пользователей), сигнализации и управления). Данная статья написана на основе материалов монографии [2], поддержанной Российским фондом фундаментальных исследований.

ЛИТЕРАТУРА

1. **Росляков А.В., Ваняшин С.В., Самсонов М.Ю. и др.** Сети следующего поколения NGN / Под ред. А.В. Рослякова. — М.: Эко-Трендз, 2008. — 424 с.
2. **Назаров А.Н., Сычев К.И.** Модели и методы расчёта показателей качества функционирования узлового оборудования и структурно-сетевых параметров сетей связи следующего поколения. — Красноярск.: Изд-во ООО «Поликом», 2010.
3. **Кокотушкин В.А., Михалёв Д.Г.** Обслуживание полнодоступным пучком нескольких потоков с относительным приоритетом на обслуживании и ограниченной общей очередью // Проблемы передачи информации. — 1969. — Вып. 2. — Т. V. — С. 53—60.
4. **Бронштейн О.И., Духовный И.М.** Модели приоритетного обслуживания в информационно-вычислительных системах. — М.: Наука, 1976. — 220 с.
5. **Зелигер Н.Б., Чугреев О.С., Яновский Г.Г.** Проектирование сетей и систем передачи дискретных сообщений. — М.: Радио и связь, 1984. — 176 с.
6. **Клейнрок Л.** Вычислительные системы с очередями. — М.: Мир, 1979. — 600 с.
7. **Боровков А.А.** Теория вероятностей. — М.: Наука, 1986. — 432 с.
8. **Norros I.** A storage model with self-similar input // Queueing Systems And Their Applications. — 1994. — № 16. — С. 387—397.
9. **Цыбаков Б.С.** Модель телетрафика на основе случайного процесса // Радиотехника. — 1999. — № 5. — С. 24—31.
10. **Шелухин О.И., Тенякшев А.М., Осин А.В.** Фрактальные процессы в телекоммуникациях / Под ред. О.И. Шелухина. — М.: Радиотехника, 2003. — 480 с.
11. **Симонина О.А.** Модели расчета показателей QoS в сетях следующего поколения: Диссертация на соискание ученой степени кандидата технических наук. — СПб.: 2005. — 129 с.

Получено после доработки 26.04.10