

УДК 621.391

АНАЛИЗ УСТОЙЧИВОСТИ И ОПТИМИЗАЦИЯ ТСР-СЕАНСОВ В МУЛЬТИСЕРВИСНЫХ ТКС



Е.В. СТАРКОВА

Харьковский национальный
университет радиоэлектроники

Запропоновано динамічну модель ТСР-сеансів, яка подана системою нелінійних диференціальних рівнянь. З використанням теорії бифуркацій отримано умови стійкості ТСР-сеансів. Запропоновано метод оптимізації ТСР-сеансів для розрахунку основних параметрів протоколу ТСР і AQM-алгоритмів.

The dynamic TCP-sessions model is proposed. The model is represented by a system of nonlinear differential equations. On the bifurcation theory basis the TCP-sessions stability conditions are obtained. A TCP-sessions optimization method for calculation of the TCP and RED-algorithm basic parameters is proposed.

Предложена динамическая модель ТСР-сеансов, представленная системой нелинейных дифференциальных уравнений. С применением теории бифуркаций получены условия устойчивости ТСР-сеансов. Предложен метод оптимизации ТСР-сеансов для расчета параметров протокола ТСР и AQM-алгоритмов.

Введение

Как показал проведенный анализ [1], современные и перспективные мультисервисные телекоммуникационные сети (ТКС) развиваются в направлении внедрения концепции сетей следующего поколения NGN (Next Generation Network). При этом в соответствии с рекомендациями Международного союза электросвязи МСЭ-Т (International Telecommunication Union, ITU-T) серии Y.2000 в основу NGN закладывается стек протоколов ТСР/IP (Transmission Control Protocol / Internet Protocol) [2].

Практика показывает, что обеспечение гарантий качества обслуживания QoS (Quality of Service) в ТКС напрямую зависит от результативности задач управления трафиком, в ходе решения которых немаловажную роль играют функции, выполняемые транспортным уровнем эталонной модели взаимодействия открытых систем (ЭМВОС). На этом уровне наиболее распространенным является протокол ТСР [2, 3], который относится к эффективным средствам решения многих задач, связанных с управлением трафиком, таких как обеспечение надежной доставки сообщений, межконцевых показателей QoS, сбалансированной загрузки сети и других.

Однако несовершенство положенных в основу ТСР математических моделей и эвристических схем нередко сопровождается необоснованным выбором численных значений основных режимов и параметров протокола. Это в свою очередь приводит к потере устойчивости ТСР-сеансов, что проявляется на практике в их колебательном режиме и медленной сходимости, и, как результат, в росте потерь пакетов и снижении качества обслуживания в ТКС в целом.

Следовательно, актуальной является научная задача, которая состоит в усовершенствовании средств управления передачей данных на транспортном уровне эталонной модели взаимодействия открытых систем путем разработки математической модели ТСП-сеансов и методов анализа их устойчивости и оптимизации, что позволит повысить производительность мультисервисных ТКС.

I. Математическая модель ТСП-сеансов с использованием AQM-алгоритмов

С целью математического описания одновременно функционирующих ТСП-сеансов с учетом классов обслуживания в соответствии с версией ТСП Tahoe [2, 3] динамику многопоточкового информационного обмена с учетом AQM-алгоритмов (Active Queue Management) отображается в виде системы уравнений:

$$\frac{d\lambda_i^k(t)}{dt} = \begin{cases} \text{режим slow start:} \\ \left(1 - P^k(t)\right) \cdot \frac{MSS}{RTT^k} \cdot \lambda_i^k(t) - P^k(t) \cdot \left(\lambda_i^k(t)\right)^2 + P^k(t) \cdot MSS \cdot \lambda_i^k(t); \\ \text{режим congestion avoidance:} \\ \left(1 - P^k(t)\right) \cdot \left(\frac{MSS}{8 \cdot RTT^k} \cdot \lambda_i^k(t) + \frac{MSS \cdot MSS}{(RTT^k)^2} \right) - \\ - P^k(t) \cdot \left(\lambda_i^k(t)\right)^2 + P^k(t) \cdot MSS \cdot \lambda_i^k(t), \end{cases} \quad (1)$$

где λ_i^k – интенсивность трафика i -го ТСП-сеанса в потоке с k -м классом обслуживания, ($i = \overline{1, M^k}$), M^k – количество ТСП-сеансов в k -м потоке, $k = \overline{1, K}$, K – количество классов обслуживания; RTT^k – время оборота пакетов k -го потока; P^k – вероятность отбрасывания (блокировки) пакетов с k -м классом обслуживания.

Вероятность отбрасывания пакетов может быть формализована в соответствии с AQM-алгоритмами, реализующими превентивное ограничение очереди до ее фактического переполнения. При этом для каждого класса обслуживания в общем случае предполагается организация отдельной очереди с различными моделями отбрасывания пакетов. Так для алгоритма произвольного раннего обнаружения перегрузки RED (Random Early Detection) [4], который получил развитие с целью учета классов обслуживания WRED (Weighted RED), расчет вероятности отбрасывания пакетов с k -м классом обслуживания производится в соответствии с выражением:

$$P^k(t) = \frac{1}{m^k} \cdot \frac{N^k(t) - N_{\min}^k}{N_{\max}^k - N_{\min}^k}, \quad (2)$$

где m^k – знаменатель граничной вероятности; N_{\min}^k , N_{\max}^k – минимальный и максимальный размер очереди соответственно; $N^k(t)$ – средний размер очереди на сетевом узле.

Для алгоритма случайной экспоненциальной маркировки REM (Random Exponential Marking) [5] выражение для вероятности P имеет следующий вид:

$$P = 1 - \phi^{-\sum_l p_l(t)}, \quad (3)$$

где $\phi > 1$ – константа; $p_l(t)$ – мера переполнения (стоимость) в l -м канале, которая определяется на основании несоответствия скорости поступающего в канал трафика и пропускной способностью этого канала, а также разности между текущим размером очереди и граничным его значением.

Для дальнейших исследований в качестве модели отбрасывания пакетов использовалось выражение (2), а средний размер очереди $N^k(t)$ рассчитывался на основании формулы Литтла [5]:

$$N^k(t) = \frac{\sum_{i=1}^{M^k} \lambda_i^k(t)}{B^k - \sum_{i=1}^{M^k} \lambda_i^k(t)}, \quad (4)$$

где B^k – пропускная способность соединения, выделенная k -му TCP-потoku.

С целью подтверждения соответствия предложенной модели (1)–(2) процессу передачи данных в реальных условиях TCP-соединения было проведено исследование одного TCP-сеанса, в ходе которого вычислялась интенсивность $\lambda(t)$. В качестве исходных данных выступали: величина доступной пропускной способности канала $B = 100$ Мбит/с; окно приема на узле-получателе, равное 64 кбайт; размер сегмента данных $MSS = 1460$ байт. Графически решения уравнений (1)–(2) при различных значениях RTT изображено на рис. 1.

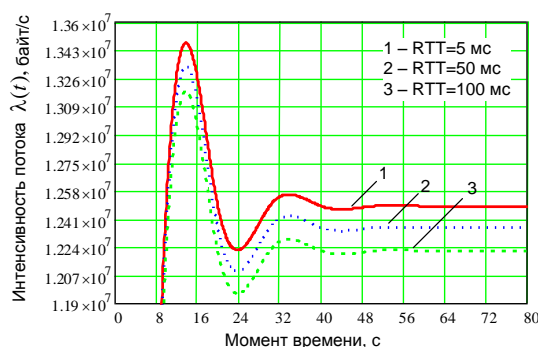


Рис. 1. Изменение интенсивности передачи данных в TCP-сеансе при различных значениях RTT

Из рис. 1 видно, что, во-первых, изменение скорости передачи $\lambda(t)$ носит колебательный характер и в некоторый момент времени устанавливается на определенном значении, что соответствует реальному процессу информационного обмена при работе протокола TCP. Во-вторых, с уменьшением времени оборота сегмента RTT (что говорит, например, об улучшении канала либо свободных канальных и буферных ресурсах) интенсивность, которую достигает TCP-поток в установившемся режиме, возрастает и стремится к значению пропускной способности соединения.

Полученная модель (1)–(2) носит четко выраженный нелинейный характер, что с математической точки зрения означает возможное наличие неединственного решения системы уравнений, более того наличие решений, являющихся неустойчивыми и приводящих к качественным изменениям поведения системы в тех или иных условиях. На практике же, как показал анализ, к потере устойчивости приводит колебание как внешних, так и внутренних параметров и условий функционирования. К внутренним относятся параметры протокола TCP, AQM-алгоритмов (в рамках предложенной модели алгоритм RED) и режимы передачи в соответствии с версией TCP. К внешним параметрам, приводящим к неустойчивости TCP-сеанса, относятся изменение структуры сети (выход из строя либо добавление сетевых каналов и узлов, что влечет изменение доступной пропускной способности), скачкообразное изменение интенсивности передаваемого трафика, увеличение задержек распространения, а также присутствие других типов трафиков (с учетом транзитных потоков и коротко живущих соединений).

Исследование реакции TCP-сеанса на колебания либо административные изменения указанных параметров и дальнейшего поведения осуществляется путем внесения изменений в исходную систему уравнений (1)–(2). Изменения касаются либо параметров, входящих в состав системы уравнений (1)–(2), либо самой структуры и вида исходных дифференциальных уравнений. На рис. 2 приведены случаи потери устойчивости, под которой подразумевается отклонение TCP-сеанса от стационарного состояния. В данном случае стационарным состоянием является режим, когда интенсивность TCP-потока со временем устанавливается на значении, близком к реальной пропускной способности соединения.

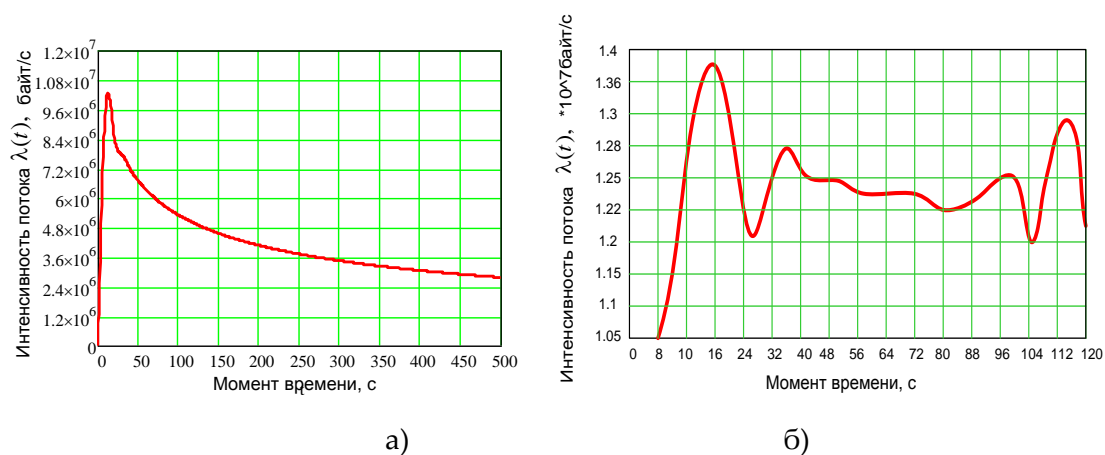


Рис. 2. Изменение интенсивности передачи данных при неустойчивом TCP-сеансе

Таким образом, стоит подчеркнуть необходимость решения такой задачи, как обнаружение и анализ причин и последствий неустойчивого функционирования TCP-сеансов. Эти исследования позволят еще на этапе математического описания ТКС, которое впоследствии будет основой для перспективных сетевых протоколов и технологий, избежать непредвиденных случаев разрыва сеансов, роста задержек, перегрузок сетевых устройств и каналов, а, следовательно, и потерь пакетов.

II. Анализ устойчивости TSP-сеансов как динамической системы

Проведенный анализ постановки и решения задачи исследования устойчивости динамических систем [7–10], к которым относится TSP-сеанс, представленных нелинейными дифференциальными уравнениями, свидетельствует о целесообразности использования возможностей теории бифуркаций и теории катастроф. Особенности применения этих теорий является возможность проанализировать динамику поведения процессов, которыми характеризуется система в окрестности ее стационарного состояния как с помощью аналитических выражений, так и с помощью графического представления (фазовых портретов).

Возможности теории бифуркаций и теории катастроф позволяют вычислить все равновесные состояния исходной системы и проанализировать влияние колебаний внутренних и внешних параметров на возможность скачкообразного перехода в то или иное состояние, которое может быть как устойчивым, так и неустойчивым. Причем с учетом особенностей процессов, протекающих в ТКС и проявляющихся в постоянных колебаниях сетевых параметров (пропускной способности, интенсивностей трафиков, структуры сети), такая возможность является немаловажной, т.к. позволяет решить задачу не только анализа влияния этих колебаний на поведение ТКС, но и синтеза, т.е. выбора структурных и функциональных параметров сети с учетом условий устойчивости. Решение такой задачи позволит избежать случаев потери устойчивости, т.е. неоправданных перегрузок сети и связанных с этим потерь пакетов данных.

Однако наиболее адекватным для анализа устойчивости представляется математический аппарат теории бифуркаций [9, 10], поскольку он основывается на моделях, описываемых дифференциальными уравнениями, и не требует необходимости построения специфических функций, как в случаях применения методов Ляпунова и теории катастроф. В рамках данной теории обеспечивается непосредственный учет параметров и переменных на уровне математического описания исходной динамической модели.

Использование математического аппарата теории бифуркаций [9, 10] для решения исходной задачи анализа и формулировки условий обеспечения устойчивости системы TSP-сеансов (1), предполагает следующая последовательность действий:

1. Поиск стационарных состояний системы дифференциальных уравнений.
2. Формирование матрицы Якоби и разложение исходных уравнений в ряд Тейлора в окрестности полученных решений (линеаризация).
3. Получение системы однородных линейных дифференциальных уравнений.
4. Вывод характеристического уравнения и нахождение его корней (собственных значений) и собственных векторов.
5. Построение траектории состояний системы (фазового пространства) и анализ поведения системы в окрестностях стационарных состояний. При этом вид траекторий состояний системы в окрестности стационарной точки (устойчивый/неустойчивый узел, седло и т.д.) определяется корнями характеристического уравнения.

Исходя из значений корней характеристического уравнения, формулируются следующие условия обеспечения устойчивости ТСР-сеансов:

- 1) для равновесного состояния типа устойчивого узла (рис. 3):

$$p_1(\lambda, B, N_{\max}) \neq p_2(\lambda, B, N_{\max}), p_1(\lambda, B, N_{\max}) < 0, p_2(\lambda, B, N_{\max}) < 0; \quad (5)$$

- 2) для равновесного состояния типа устойчивого фокуса (рис. 4):

$$p_{1,2}(\lambda, B, N_{\max}) = \alpha \pm i\beta, \alpha > 0. \quad (6)$$

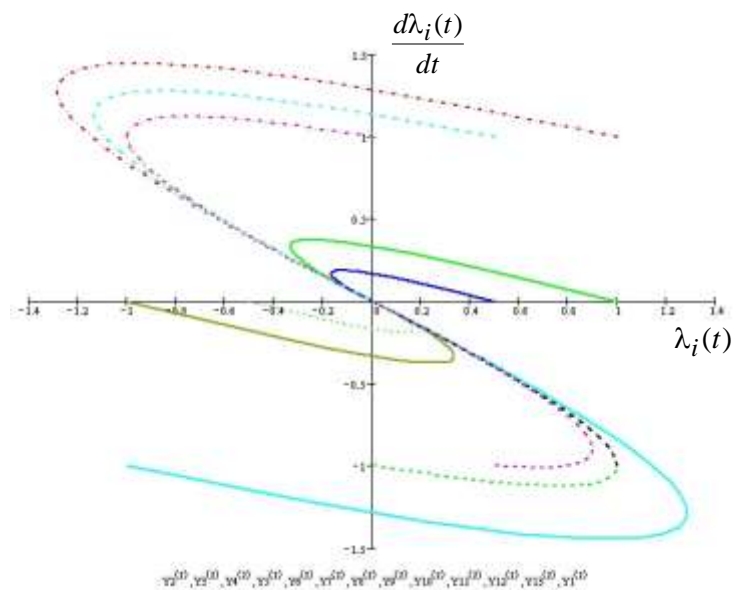


Рис. 3. Фазовый портрет системы ТСР-сеансов: равновесная точка типа устойчивый узел

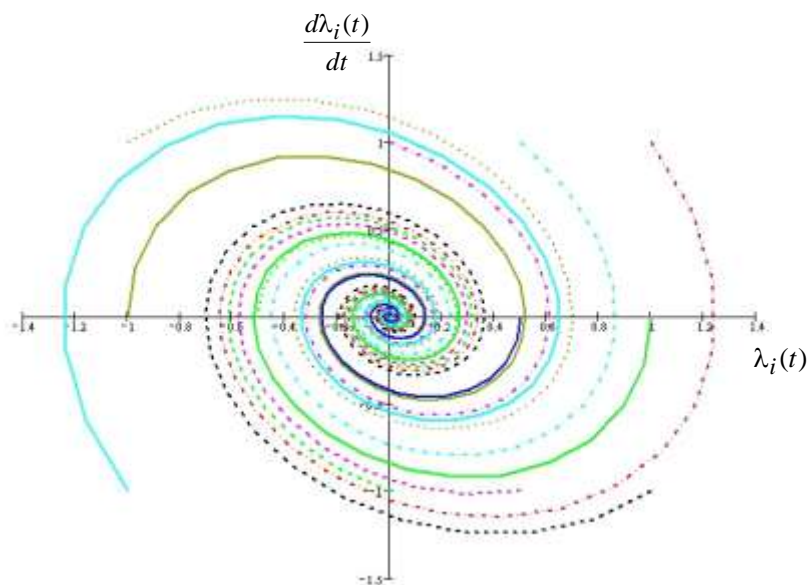


Рис. 4. Фазовый портрет системы ТСР-сеансов: равновесная точка типа устойчивый фокус

Таким образом, предложенный метод анализа устойчивости позволил исследовать динамическую модель TCP-сеансов, представленную системой нелинейных дифференциальных уравнений (1). В ходе решения этой задачи в аналитическом виде были получены условия обеспечения устойчивости TCP-сеансов (5)–(6), которые можно также построить и проанализировать геометрически в фазопараметрическом пространстве (рис. 3, 4). Причем на характер решений (стационарных состояний и соответствующего поведения системы в окрестности этих состояний) влияют структурные и функциональные сетевые параметры (параметры TCP/AQM-алгоритмов, положенных в основу исходной модели (1)–(2), а также топология сети, колебания интенсивности передаваемого трафика, многопоточность). Полученные условия обеспечения устойчивости TCP-сеансов могут быть использованы с целью решения задачи обоснованного выбора параметров TCP/AQM-алгоритмов в ходе дальнейшей оптимизации TCP-сеансов.

III. Оптимизация процесса передачи данных в рамках TCP-сеансов

С целью обоснованного выбора параметров протокола TCP и AQM-алгоритмов (RED/WRED) необходимо сформулировать и решить оптимизационную задачу, связанную с минимизацией (максимизацией) некоторого целевого функционала потерь (качества). С целью моделирования интенсивности потерь введем величину $\lambda_i^{k\text{отк}}(t)$, которая характеризует интенсивность отказов в обслуживании i -му TCP-сеансу в потоке с k -м классом обслуживания. Тогда для $\lambda_i^{k\text{отк}}(t)$ справедливо соотношение:

$$\lambda_i^{k\text{отк}}(t) = \lambda_i^{k\text{прб}}(t) \cdot \lambda_i^k(t), \quad (7)$$

где $\lambda_i^{k\text{прб}}(t)$ – QoS-требования относительно интенсивности i -го TCP-сеанса в установленном (стационарном) состоянии, которое достигается при сбалансированном использовании доступной пропускной способности канала; τ – временной интервал, который, как правило, соизмерим со средней длительностью TCP-сеанса в сети.

В современных мультисервисных ТКС важным требованием является предоставление гарантированной пропускной способности определенным потокам между источником и получателем. Такая процедура организуется с помощью дифференциации TCP-сеансов посредством назначения приоритетов и выделения заданной величины пропускной способности в соответствии с запрашиваемыми требованиями. Причем поток с наивысшим приоритетом, может обслуживаться в первую очередь, тогда как остальные потоки получают отказ в обслуживании, как в алгоритмах PQ (Priority Queuing) и LLQ (Low Latency Queuing) [4]. Может быть организовано более справедливое обслуживание потоков, когда потокам с более низкими приоритетами все же предоставляется некоторая доля пропускной способности даже при обслуживании высокоприоритетного потока, как в алгоритме CBWFQ (Class-Based Distributed Weighted Fair Queuing) [4].

С точки зрения поиска компромисса между необходимостью учета технологических особенностей процесса передачи данных в рамках TCP-сеансов, с одной сто-

роны, и возможностью получения численных результатов расчета в рамках известных оптимизационных процедур, с другой, воспользуемся критерием, определяющим минимум функционала:

$$J = \int_0^{\tau} \left[\sum_{k=1}^K \sum_{i=1}^{M^k} \alpha_i^k \cdot \lambda_i^{k\text{отк}}(t) \right] dt, \quad (8)$$

где α_i^k – приоритет i -го ТСП-сеанса.

Использование критерия (8) гарантирует минимизацию суммарной взвешенной относительно приоритета интенсивности отказов в обслуживании на транспортном уровне. При этом реализуется схема распределения пропускной способности между потоками, когда каждому потоку предоставляется четко заданная величина пропускной способности, назначаемая административно.

После преобразований выражение (7) можно привести к виду:

$$\max_{\lambda_i^k(t)} J = \max_{\lambda_i^k(t)} \int_0^{\tau} \left[\sum_{k=1}^K \sum_{i=1}^{M^k} \frac{\alpha_i^k}{\tau} \cdot \lambda_i^k(t) \right] dt. \quad (9)$$

Физический смысл функционала (9) теперь уже состоит в максимизации суммарной взвешенной относительно приоритета интенсивности ТСП-потоков на интервале τ . Такая постановка задачи гарантирует выполнение важного свойства системы ТСП-сеансов – эффективности, под которой в соответствии с принципами функционирования протокола ТСП подразумевается передача с максимальной суммарной интенсивностью в пределах доступной пропускной способности всех потоков данных.

При решении задачи оптимизации в качестве управляемых выступают параметры протокола ТСП (величина MSS, время RTT, размер окна перегрузки, значение порога, режимов передачи), а также параметры алгоритмов борьбы с перегрузками RED/WRED (величины N_{\min} , N_{\max} , m). На практике можно оптимизировать только часть из перечисленных параметров, при этом другие характеристики фиксировать на некоторых значениях, например, в соответствии с настройками протоколов по умолчанию.

Минимизируемый функционал, сформулированный в скалярном уравнении (8), удобно представить в векторно-матричной форме:

$$J = \int_0^{\tau} \sum_{k=1}^K A_k^T \Lambda_{k\text{отк}}(t) dt, \quad (10)$$

где $A_k = [\alpha_1^k, \alpha_2^k, \dots, \alpha_i^k, \dots, \alpha_N^k]^T$ – вектор приоритетов для ТСП-потоков;

$\Lambda_{k\text{отк}} = [\lambda_1^{k\text{отк}}, \lambda_2^{k\text{отк}}, \dots, \lambda_i^{k\text{отк}}, \dots, \lambda_N^{k\text{отк}}]^T$ – вектор, элементами которого являются интенсивности отказов в обслуживании ТСП-сеансов $\lambda_i^{k\text{отк}}$, рассчитываемых в соответствии с выражением (7).

Стоит отметить, что исследование и дальнейшее обеспечение устойчивости процессов информационного обмена и управления ресурсами сети, протекающих в

реальных ТКС, существенно влияет на эффективность их работы. Это позволяет выявить и предупредить внезапное переполнение очередей сетевых узлов и каналов связи, разрывы сеансов связи и другие факторы, приводящие к неустойчивости ТСП-сеансов при незначительных изменениях структурных и функциональных параметров. Следовательно, в ходе решения оптимизационной задачи, кроме выполнения условия, связанного с динамикой ТСП-сеансов (1)–(2), необходимо учитывать полученные ранее условия обеспечения устойчивости системы ТСП-сеансов (5) и (6).

Для обеспечения более справедливого обслуживания ТСП-потоков, при котором даже в условиях перегрузок низкоприоритетному трафику гарантируется некоторая доля пропускной способности, предлагается использовать квадратичный целевой функционал с теми же ограничениями в виде исходной модели (1)–(2) и условий обеспечения устойчивости системы ТСП-сеансов (5) и (6). В таком случае целевой функционал, представленный выражением (10), принимает вид:

$$J = \int_0^{\tau} \sum_{k=1}^K \Lambda_{k\text{отк}}^T(t) H_k \Lambda_{k\text{отк}}(t) dt, \quad (11)$$

где H – диагональная матрица, имеющая структуру

$$H = \begin{pmatrix} \alpha_1^k & 0 & \dots & 0 \\ 0 & \alpha_2^k & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \alpha_N^k \end{pmatrix}, \quad (12)$$

координаты которой определяют величину приоритетов, согласно которым осуществляется обслуживание и ограничение трафиков ТСП-сеансов.

С целью сравнения результатов решения оптимизационной задачи с различными целевыми функциями приоритеты a_1 и a_2 ТСП-потоков задавались в пределах от 0 до 7 (0 – наивысший, 7 – наименьший приоритеты) следующим образом:

- первый случай, когда у первого (высокоприоритетного) потока значение приоритета равняется $a_1 = 7$, у второго (низкоприоритетного) – $a_2 = 0$;
- второй случай – значения приоритетов $a_1 = 5$ и $a_2 = 2$ для первого и второго ТСП-потока соответственно;
- третий случай – приоритеты ТСП-потоков одинаковые и равны $a_1 = a_2 = 4$.

На рис. 5 и 6 представлены графики, отображающие интенсивности отказов высокоприоритетного $\lambda_1^{\text{отк}}$ и низкоприоритетного $\lambda_2^{\text{отк}}$ потоков в зависимости от их интенсивностей передачи λ_1 и λ_2 для случая использования линейного целевого функционала. На рис. 7 и 8 показаны те же зависимости для случая использования квадратичного целевого функционала.

Исходя из того, что пропускная способность соединения между оконечными узлами равна 120 Кбит/с, интенсивности первого и второго ТСП-потоков изменялись от нуля до 200 Кбит/с. Полученные зависимости (рис. 5) показывают, что первый (высокоприоритетный) ТСП-поток получал отказ в обслуживании при $\lambda_1 = 120$ Кбит/с

(рис. 5, а) только в том случае, когда суммарной пропускной способности ТСП-соединений было недостаточно для обслуживания обоих потоков, а менее приоритетный ТСП-поток был полностью ограничен (рис. 5, б). Величина отказов второго ТСП-потока увеличивалась с ростом как собственной интенсивности, так и интенсивности более приоритетного потока (рис. 5, б).

Стоит отметить, что при изменении приоритетов величины отказов для высокоприоритетного и низкоприоритетного ТСП-потоков практически не изменились, то есть с увеличением интенсивности пакеты первого потока ($a_1 = 5$) начинали отбрасываться только тогда, когда все пакеты второго потока ($a_2 = 2$) были полностью отброшены. И только в случае одинаковых приоритетов ($a_1 = a_2 = 4$) ограничение ТСП-потоков с увеличением их интенсивности происходило одинаково (рис. 6). Только в таком случае осуществлялось справедливое распределение пропускной способности между ТСП-сеансами, и ни один из потоков не получал полного отказа.

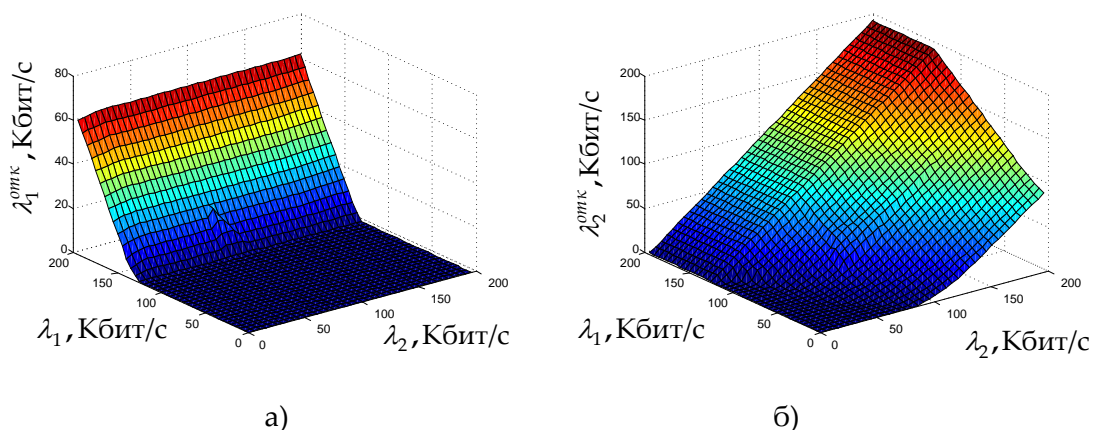


Рис. 5. Зависимость интенсивности отказов высокоприоритетного $\lambda_1^{отк}$ ($a_1 = 7$) и низкоприоритетного $\lambda_2^{отк}$ ($a_2 = 1$) ТСП-потоков от их интенсивностей передачи λ_1 и λ_2 при использовании критерия (10)

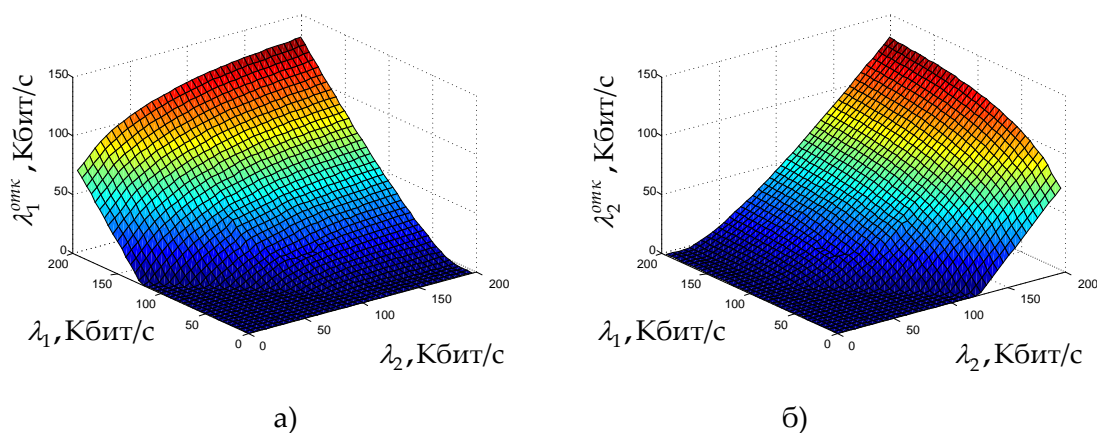


Рис. 6. Зависимость интенсивности отказов высокоприоритетного $\lambda_1^{отк}$ и низкоприоритетного $\lambda_2^{отк}$ ТСП-потоков при $a_1 = a_2 = 4$ от их интенсивностей передачи λ_1 и λ_2 при использовании критерия (10)

Как показывают результаты исследования предложенного метода оптимизации ТСП-сеансов на основе использования квадратичного целевого функционала (11) с ограничениями (1)–(2), несмотря на увеличение интенсивности потоков, низкоприоритетный поток не отбрасывается полностью. При изменении приоритетов трафиков распределение доступной пропускной способности изменяется и, соответственно, изменяется доля отказов. Так при интенсивности передачи ТСП-потоков, равной $\lambda_1 = \lambda_2 = 75$ Кбит/с интенсивность отказов высокоприоритетного трафика ($a_1 = 7$) составляет $\lambda_1^{отк} = 5$ Кбит/с (рис. 7, а), а низкоприоритетного ($a_2 = 0$) $\lambda_2^{отк} = 22$ Кбит/с (рис. 7, б). В случае снижения приоритета первому трафику (например, $a_1 = 5$) и повышения приоритета второму трафику (например, $a_2 = 2$) при тех же интенсивностях передачи интенсивности отказов составляют $\lambda_1^{отк} = 9$ Кбит/с и $\lambda_2^{отк} = 22$ Кбит/с. Одинаковые приоритеты ($a_1 = a_2 = 4$) дают идентичную интенсивность отказов, т.е. $\lambda_1^{отк} = \lambda_2^{отк} = 11$ Кбит/с при интенсивности передачи $\lambda_1 = \lambda_2 = 75$ Кбит/с (рис. 8).

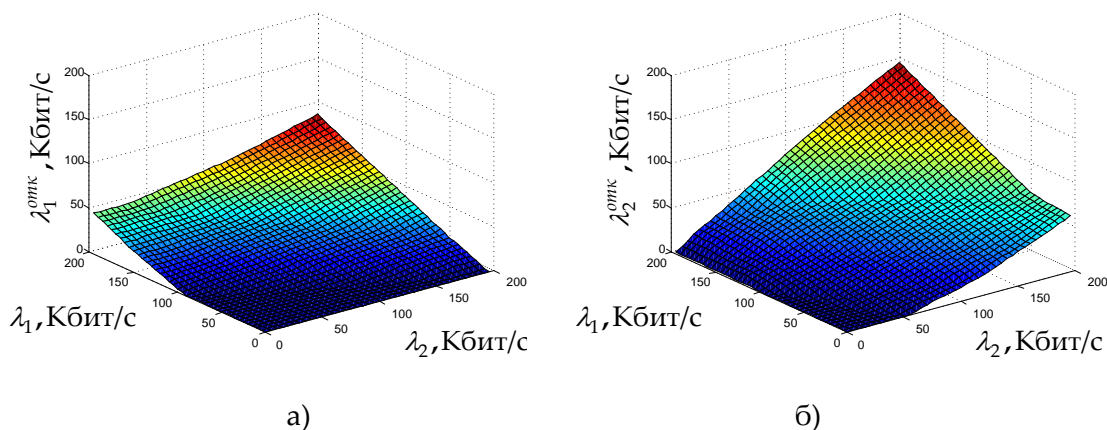


Рис. 7. Зависимость интенсивности отказов высокоприоритетного $\lambda_1^{отк}$ ($a_1 = 7$) и низкоприоритетного $\lambda_2^{отк}$ ($a_2 = 1$) ТСП-потоков от их интенсивностей передачи λ_1 и λ_2 при использовании критерия (11)

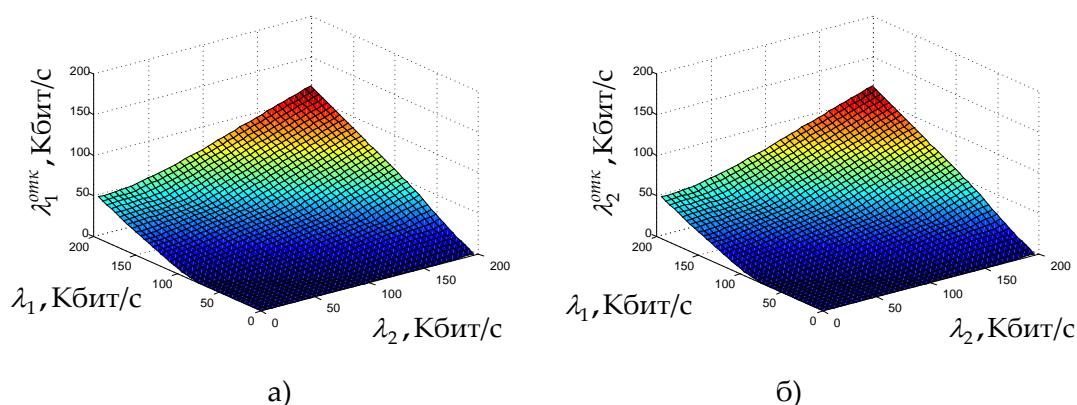


Рис. 8. Зависимость интенсивности отказов высокоприоритетного $\lambda_1^{отк}$ и низкоприоритетного $\lambda_2^{отк}$ ТСП-потоков при $a_1 = a_2 = 4$ от их интенсивностей передачи λ_1 и λ_2 при использовании критерия (11)

Как показывают результаты моделирования, при интенсивностях ТСП-потоков $\lambda_1 = \lambda_2 = 200$ Кбит/с, что в сумме превышает пропускную способность соединения (120 кбит/с), в случае использования линейного функционала (10) интенсивность отказов первого потока составляет $\lambda_1^{omk} = 60$ Кбит/с ($a_1 = 7$) (рис. 5, а) при полном ограничении второго $\lambda_2^{omk} = 200$ Кбит/с ($a_2 = 1$) (рис. 5, б). В случае использования квадратичного функционала (11) интенсивность отказов первого ТСП-потока составляет $\lambda_1^{omk} = 89$ Кбит/с (рис. 7, а), а второго $\lambda_2^{omk} = 148$ Кбит/с (рис. 7, б). ТСП-потоки с приоритетами $a_1 = 5$ и $a_2 = 2$ в условиях перегрузки имеют следующие интенсивности отказов: при использовании линейного функционала $\lambda_1^{omk} = 70$ Кбит/с и $\lambda_2^{omk} = 200$ Кбит/с, при использовании квадратичного функционала $\lambda_1^{omk} = 105$ Кбит/с и $\lambda_2^{omk} = 132$ Кбит/с.

Таким образом, использование квадратичного функционала (11) в ходе решения оптимизационной задачи позволяет уменьшить интенсивность отказов низкоприоритетного ТСП-трафика за счет более рационального распределения доступной пропускной способности между высокоприоритетным и низкоприоритетным потоками, чем в случае использования линейного функционала (10). Такой результат становится особенно очевидным в условиях перегрузки сети, когда суммарная интенсивность передаваемых потоков превышает доступную пропускную способность соединения. Наиболее приемлемым представляется использование нелинейного программирования [11, 12] для решения поставленной оптимизационной задачи, а сама реализация метода оптимизации процесса передачи данных в рамках ТСП-сеансов предполагает следующую последовательность действий:

1. Мониторинг и сбор исходных данных о состоянии ТСП-сеансов, режимах передачи в протоколе ТСП, параметров AQM-алгоритмов (RED/WRED) на узлах ТКС.
2. На основе анализа собранных исходных данных о состоянии сети устанавливается их аналитическая взаимосвязь путем анализа математической модели (1)–(2).
3. Формализация условий обеспечения устойчивости ТСП-сеансов (5) и (6).
4. Постановка оптимизационной задачи (8), в ходе решения которой обеспечивается минимизация отказов в обслуживании трафиков пользователей, и согласованно рассчитываются основные параметры протокола ТСП и AQM-алгоритмов.
5. Параметры протокола ТСП и AQM-алгоритмов в автоматическом режиме подлежат корректировке (настройке), обеспечивая тем самым оптимизацию ТСП-сеансов в целом.
6. По истечении интервала RTT осуществляется повторная постановка и решение оптимизационной задачи (8), т.е. производится переход к п.1.

В некоторых случаях периодический характер расчетов может быть нарушен, например, при «нештатном» изменении состояния ТКС и ее ТСП-соединений, вызванном изменением структуры сети, скачкообразном изменении интенсивности поступающего в сеть трафика или числа ТСП-потоков.

Выводы

Анализ состояния и перспектив развития современных ТКС показал, что ключевыми аспектами при построении сетей NGN является их способность обслуживать разнородный по типу и требованиям трафик, гарантируя соответствие предоставленного уровня качества обслуживания запрашиваемому пользователями. Однако большинство средств управления трафиком, среди которых ТСП/AQM-алгоритмы, не отвечают требованиям по обеспечению QoS и сами нередко становятся причинами потери устойчивости ТКС. Установлено, что это обусловлено недостатками математических моделей и эвристических схем, положенных в основу этих алгоритмов.

В статье предложена математическая модель ТСП-сеансов, основанная на аппарате дифференциальных уравнений и теории массового обслуживания и представлена в пространстве состояний. К характерным свойствам предложенной модели относится учет динамичности процессов информационного обмена, а также изменения режимов передачи данных протокола ТСП. Также важным является возможность описания этих режимов согласно существующим и перспективным версиям протокола ТСП. Отличительной чертой модели является учет особенностей передачи одновременно нескольких ТСП-потоков с разными классами обслуживания, что свойственно реальным мультисервисным ТКС. Кроме того, в рамках единой модели (1)–(2) формализованы как процессы передачи данных в соответствии с протоколом ТСП, так и процессы борьбы с перегрузками (AQM), что позволяет конечным устройствам оперативно реагировать на вероятные перегрузки в сети и, соответственно, избегать необоснованных потерь пакетов. Преимуществом предложенной модели ТСП-сеансов является возможность выбора типа AQM-алгоритма за счет использования различных моделей блокировки (отбрасывания) пакетов. Предложенная модель (1)–(2) в дальнейшем может быть использована для анализа устойчивости ТСП-сеансов и оптимизации выбора параметров ТСП/AQM-алгоритмов. Причем важность решения задачи обеспечения устойчивости продиктована необходимостью учета колебаний параметров сети (параметров ТСП/AQM-алгоритмов, флуктуаций трафика), особенно в условиях, близких к перегрузкам.

Предложен метод анализа устойчивости ТСП-сеансов в рамках динамической модели (1)–(2), основанный на теории бифуркаций. В отличие от классических методов математического анализа, позволяющих исследовать плавные непрерывные процессы, теория бифуркаций является универсальным инструментарием для исследования скачкообразных переходов, разрывов, внезапных качественных изменений. Использование данной теории при математическом описании ТКС открывает широкие возможности по обеспечению структурной и функциональной устойчивости по отношению к резким непредвиденным изменениям ее структуры (например, при выходе из строя сетевого элемента или целого участка сети), а также условиям функционирования (например, при скачкообразном увеличении интенсивности трафика, поступающего в сеть или изменении пропускной способности каналов связи). Метод основан на определении стационарных состояний (решений) системы (1) и последующем анализе поведения ТСП-сеансов в окрестности этих состояний, среди

которых наиболее важными являются устойчивый узел и устойчивый фокус. Применение предложенного метода позволило получить в аналитическом виде условия обеспечения устойчивости ТСП-сеансов по отношению к незначительным колебаниям ТСП/RED-параметров.

Предложен метод оптимизации ТСП-сеансов, в рамках которого обеспечивается расчет основных параметров протокола ТСП и AQM-алгоритмов (RED/WRED), что позволило с учетом условий устойчивости (5) и (6) минимизировать суммарную взвешенную относительно приоритетов интенсивность отказов в обслуживании различных типов трафика (максимизировать суммарную взвешенную относительно приоритетов интенсивность ТСП-поток). Использование метода предусматривает мониторинг и сбор исходных данных о состоянии ТСП-сеансов, установление аналитических зависимостей между параметрами ТСП и RED с использованием математической модели (1)–(2); формализацию условий обеспечения устойчивости ТСП-сеансов (5) и (6); постановку и решение оптимизационной задачи (8); расчет параметров протокола ТСП и AQM-алгоритмов (RED/WRED) и их последующую автоматическую коррекцию.

Список литературы:

1. Бакланов И. Г. NGN: принципы построения и организации. Под ред. Ю. Н. Чернышова. – М.: Эко-Трендз, 2008. – 400 с.
2. Стивенс У. Р. Протоколы ТСП/IP. Практическое руководство: пер. с англ. и коммент. А.Ю. Глебовского. – СПб.: «Невский диалект»-«БХВ-Петербург», 2003. – 672 с.
3. Postel J. Transmission control protocol. RFC 793. – California, sept. 1981. – 85 p.
4. Вегешна III. Качество обслуживания в сетях IP: пер. с англ. – М.: Изд. дом «Вильямс», 2003. – 368 с.
5. Sanjeeva Athuraliya, Victor H. Li, Steven H. Low, Qinghe Yin REM: Active Queue Management // IEEE Network. – 2001. – P. 121–132.
6. Крылов В.В., Самохвалова С.С. Теория телетрафика и ее приложения – СПб.: БХВ-Петербург, 2005. – 288 с.
7. Гёльднер К., Кубик С. Нелинейные системы управления: пер. с нем. – М.: Мир, 1987. – 368 с.
8. Арнольд В.И. Теория катастроф – [3-е изд., доп.]. – М.: Наука. Гл. ред. физ.-мат. лит., 1990. – 128 с.
9. Теория бифуркаций / Арнольд В. И., Афраймовш В. С., Ильяшенко Ю. С., Шильников Л. П.. – М: ВИНТИ. Том 5. – 1986. – 218 с.
10. Анищенко В.С. Устойчивость, бифуркации, катастрофы // Соросовский образовательный журнал. – 2000. – № 6. – С. 105–109.
11. Афанасьев В.Н., Колмановский В.Б., Носов В.Р. Математическая теория конструирования систем управления: Учеб. пособие для вузов. – М.: Высш. шк., 1989. – 447 с.
12. Справочник по теории автоматического управления / под ред. А.А. Красовского. – М.: Наука. – 1987. – 712 с.