

УДК 621.391

А.В. Лемешко, Е.В. Старкова

Харьковский национальный университет радиоэлектроники, Харьков

МЕТОД ОПТИМИЗАЦИИ ПРОЦЕССА ПЕРЕДАЧИ ДАННЫХ В РАМКАХ TCP-СЕАНСОВ

В работе проведен анализ существующих версий протокола TCP, а также перспективных решений в области управления передачей пакетов на транспортном уровне Эталонной модели взаимодействия открытых систем (ЭМВОС). Выявлен ряд недостатков математических моделей и методов, положенных в основу данного протокола, устранение которых связано с оптимизацией соответствующих сетевых параметров. Предложен подход к решению задачи оптимизации, основанный на динамической модели TCP-сеансов и учитывающий ряд условий-ограничений, направленных на обеспечение заданных показателей качества и устойчивости TCP-сеансов.

Ключевые слова: TCP-сеанс, процесс передачи данных.

Введение

Как показал анализ, современные и перспективные мультисервисные телекоммуникационные сети (ТКС) развиваются в направлении внедрения концепции сетей следующего поколения NGN (Next Generation Network) [1, 2]. Эффективность ТКС во многом зависит от результативности постановки и решения задач, связанных с управлением сетевыми ресурсами. В основу NGN в соответствии с рекомендациями Международного союза электросвязи серии Y 2000 закладывается стек протоколов TCP/IP [3], в рамках которого немаловажную роль играют функции, выполняемые транспортным уровнем ЭМВОС. В частности, основными задачами протокола TCP (Transmission Control Protocol) является гарантия успешной доставки данных и обеспечение обмена данными между процессами, выполняющимися на оконечных системах (устройствах, узлах), при помощи службы обмена данными, предоставляемой протоколом сетевого уровня [4]. Протокол TCP является эффективным средством управления потоками данных, сбалансированной загрузки сети, борьбы с перегрузками, обеспечения сквозных показателей качества и др. Однако его эффективность во многом зависит от корректности выбора тех или иных параметров как самого TCP протокола, так и других взаимодействующих с ним средств борьбы с перегрузкой в ТКС.

Анализ существующих решений. Анализ моделей и методов, положенных в основу базового протокола TCP и его множественных версий (TCP Reno, TCP Vegas, TCP Tahoe, TCP Westwood, Scalable TCP и т.д.) [3, 4], показал ряд присущих им недостатков, с целью минимизации которых сформулированы следующие требования, предъявляемые к перспективным решениям:

1) учет динамичности процессов передачи данных на транспортном уровне ТКС;

2) учет особенностей передачи одновременно нескольких TCP-потоков, что свойственно для реальных мультисервисных ТКС, использующих для передачи данных протокол TCP (многие протоколы прикладного уровня, например, FTP, Telnet, HTTP, POP3, SMTP и др. работают именно с этим протоколом);

3) учет изменения режимов передачи данных в рамках процесса функционирования протокола TCP;

4) обеспечение согласованного использования механизмов борьбы с перегрузками и протокола транспортного уровня TCP, что позволяет оперативно реагировать на вероятные перегрузки в сети и соответственно, избежать необоснованных потерь пакетов;

5) ориентация на различные версии TCP, что должно придать гибкость в использовании той или иной существующей реализации данного протокола, а также ориентировать на использование в перспективных решениях в этой области;

6) возможность выбора различных моделей блокировки (отбрасывания) пакетов в механизмах борьбы с перегрузками, что позволит адаптироваться под структурные и функциональные особенности сети и характеристики передаваемых трафиков, своевременно реагировать на изменения этих параметров;

7) обеспечение устойчивости TCP-сеансов за счет формулировки соответствующих условий в виде дополнительных ограничений в рамках существующих моделей и методов сетевого управления либо как некоторой целевой функции.

Следовательно, актуальной научной задачей является оптимизация TCP-сеансов в мультисервисных ТКС, что связано с необходимостью обоснованного выбора численных значений параметров протокола TCP и параметров взаимодействующих с ним средств борьбы с перегрузками, таких как RED, WRED и т.д.

Математическая модель TCP-сеансов

Следует учесть, что в целом для математического описания ТКС на сегодняшний день существует множество подходов, в основу которых положено использование возможностей различных математических аппаратов [5]. Среди них можно выделить теоретико-игровые модели, модели принятия оптимальных решений, модели управляющих автоматов, вероятностные и агрегативные модели сложных систем. Каждый из математических аппаратов определяет свой подход к формализации задач, связанных с процессами маршрутизации, управления сетевыми параметрами и процессами информационного обмена, которые, в конечном счете, формулируются как оптимизационные задачи.

Наряду с перечисленными математическими подходами, заслуживает внимания аппарат дифференциально-разностных уравнений состояния сети, который достаточно точно соответствует предъявляемым требованиям. В рамках модели пространства состояний удовлетворяются требования по учету динамичности процессов (в том числе и TCP-сеанса), протекающих в ТКС, изменение сетевых параметров (структурных и функциональных), а также учету возможности одновременного существования нескольких потоков в сети.

Рассмотрим процесс информационного обмена в IP-сети, где при установке и поддержке TCP-сеанса с целью предотвращения возможных перегрузок на промежуточных узлах связи и последующих потерь пакетов используется дополнительные алгоритмы RED/WRED (Random Early Detection/Weighted Random Early Detection).

Анализ различных версий данного протокола показал, что в процессе работы TCP можно выделить несколько стандартных режимов (фаз) передачи данных (Slow Start, Congestion Avoidance, Fast Recovery) [3, 4]. Переход в тот или иной режим, начиная с установления соединения, зависит от степени загруженности каналов связи, буферов промежуточных и приемного устройств, количества ошибок и искажений данных. Тогда модель TCP-сеансов в пространстве состояний можно представить системой неавтономных дифференциальных уравнений, отражающей динамику передачи данных между узлом-отправителем и узлом-получателем при наличии в канале нескольких TCP-потоков:

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

где $\lambda_i(t)$ – интенсивность TCP-потока, передаваемого в i -м сеансе, $i = \overline{1, M}$; MSS – максимальный размер сегмента (Maximum Segment Size); RTT – период кругового обращения пакета (Round Trip Time); $P(t)$ – вероятность блокировки (отбрасывания) пакетов.

В рамках предложенной модели вероятность блокировки (отбрасывания) пакетов $P(t)$ может представляться произвольной (но обоснованной) функцией. В частном случае, при использовании в качестве механизма ограничения длины очереди алгоритма RED, вероятность блокировки определяется согласно равенства [6]:

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

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

Очередь на сетевом узле (маршрутизаторе) включает пакеты суммарного потока, что позволяет учесть деление пропускной способности между TCP-сеансами в пределах одного канала. В свою очередь, система (1) дает возможность применения к различным TCP-потокам приоритизации на основе выделения разной пропускной способности, как это реализуется, например, в дисциплине обслуживания очередей WRED [6]. В механизме WRED, в частности предполагается использование отдельных значений граничной вероятности и весов для разных IP-приоритетов, позволяя обеспечивать различные уровни качества обслуживания для различных типов трафика в периоды переполнения очередей сетевых узлов и каналов.

В модели предполагается, что средний размер очереди на сетевом узле можно рассчитать согласно положениям теории массового обслуживания, исходя из формулы Литтла [7, 8]:

$$N(t) = \frac{\lambda(t)}{B - \lambda(t)} - \frac{\lambda(t)}{B}, \quad (3)$$

где B – пропускная способность канала связи.

Метод оптимизации процесса передачи данных в рамках TCP-сеансов

В основу предлагаемого метода, связанного с оптимизацией выбора параметров протокола TCP и алгоритмов RED/WRED, положено решение оптимизационной задачи, в ходе которой осуществляется минимизация (максимизация) некоторого целевого функционала потерь (качества). С точки зрения поиска компромисса между необходимостью учета физики процессов передачи данных в рамках TCP-сеансов, с одной стороны, и возможностью получения численных результатов расчета в рамках из-

вестных оптимизационных процедур, с другой стороны, воспользуемся критерием, определяющим минимум функционала:

$$J = \min \left[\sum_{i=1}^M a_i \cdot \left(\lambda_i^{\text{треб}} - \frac{1}{T} \cdot \int_0^T \lambda_i(t) dt \right) \right], \quad (4)$$

где a_i – приоритет i -го ТСП-потока; T – временной интервал оптимизации, который, как правило, соизмеримым со средней длительностью ТСП-сеанса в сети.

Использование критерия (4) гарантирует минимизацию взвешенных относительно приоритета отказов в обслуживании на транспортном уровне. После преобразований выражение (4) можно привести к виду:

$$J = \sum_{i=1}^M (a_i \cdot \lambda_i^{\text{треб}}) - \min \left[\sum_{i=1}^M \left(\frac{a_i}{T} \cdot \int_0^T \lambda_i(t) dt \right) \right] = \max \left[\sum_{i=1}^M \frac{a_i}{T} \cdot \int_0^T \lambda_i(t) dt \right]. \quad (5)$$

Физический смысл функционала (5) теперь уже состоит в максимизации средней суммарно взвешенной относительно приоритета интенсивности ТСП-потоков на интервале T .

При решении задачи оптимизации в качестве управляемых выступают параметры протокола TCP (величина MSS, интервал 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.

Учитывая мультисервисный характер современных ТКС в ходе решения оптимизационной задачи, кроме выполнения условия, связанного с динамикой TCP-сеансов, необходимо учитывать требования к показателям качества обслуживания (Quality of Service, QoS). Количественная оценка показателей QoS может быть определена, например, на основании рекомендаций Y 1540, Y 1541 Международного союза электросвязи.

С целью формулировки условий обеспечения гарантированного качества обслуживания введены ограничения по временным показателям QoS:

$$\sum_{j=1}^K \tau_i^j(t) \leq \tau_i^{\text{треб}}; \quad (6) \quad \sum_{j=1}^K \sigma_i^j(t) \leq \sigma_i^{\text{треб}} \quad (7)$$

и по показателям надежности:

$$\prod_{j=1}^K P_i^j(t) \leq P_i^{\text{треб}}, \quad (8)$$

где параметр K определяет количество участков сети между узлом-источником и узлом-получателем; $\tau_i^j(t)$, $\sigma_i^j(t)$, $P_i^j(t)$ – текущие значения средней задержки, джиттера и вероятности по-

терь пакетов в j -м канале связи TCP-соединения для i -го ТСП-потока; $\tau_i^{\text{треб}}$, $\sigma_i^{\text{треб}}$ и $P_i^{\text{треб}}$ – требования, предъявляемые к межконцевым значениям тех же показателей QoS.

При этом в выражении (6) формализовано требование к межконцевой (суммарной по всем участкам сети между узлом-источником и узлом-получателем) средней задержке пакетов в сети, в выражении (7) – к джиттеру, а соотношение (8) отображает требование относительно величины вероятности отбрасывания (блокировки) пакетов.

Стоит отметить, что исследование и дальнейшее обеспечение устойчивости процессов информационного обмена и управления сетевыми ресурсами, 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 TCP-сеансов при незначительных изменениях структурных и функциональных параметров. Анализ известных подходов к анализу устойчивости динамических систем, к которым относится TCP-сеанс, описываемых нелинейными дифференциальными уравнениями, позволил сделать вывод в пользу применения возможностей теории бифуркаций [9, 10]. В рамках решения оптимизационной задачи на основании данной теории и модели (1) ряд ограничений (6) – (8) может быть дополнен условиями обеспечения устойчивости, полученными в [11].

Таким образом, реализация метода оптимизации процесса передачи данных в рамках TCP-сеансов предполагает следующую последовательность действий:

1. Мониторинг и сбор исходных данных о текущих значениях TCP-соединения, режиме передачи в протоколе TCP, параметров алгоритмов предотвращения перегрузки (RED/WRED) на узлах анализируемой сети.

2. На основе анализа собранных исходных данных о состоянии сети устанавливается их аналитическая взаимосвязь путем анализа аналитической модели (1).

3. Формализация условий-ограничений по временным показателям QoS (6) и (7), по показателям надежности (8), а также условий обеспечения устойчивости TCP-сеансов.

4. Постановка оптимизационной задачи (5), в ходе решения которой обеспечивается минимизация отказов в обслуживании трафиков пользователей и согласованно рассчитываются основные параметры протокола TCP и алгоритмов борьбы с перегрузкой, например, алгоритмов RED/WRED.

5. Параметры протокола TCP и алгоритмов борьбы с перегрузкой, например, алгоритмов RED/WRED в автоматическом режиме подлежат корректировке (настройке), обеспечивая, тем самым,

оптимізацію TCP-сеансів в цілому.

6. По истечении интервала RTT осуществляется повторная постановка и решение оптимизационной задачи (5), т.е. производится переход к п.1.

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

Выводы

Таким образом, с целью оптимизации параметров протокола TCP и взаимодействующих с ним механизмов борьбы с перегрузкой предложен соответствующий метод, связанный с представлением TCP-сеансов системой нелинейных дифференциальных уравнений. В основу метода положено решение сформулированной на выходе нелинейной дифференциальной модели TCP-сеансов оптимизационной задачи, где в качестве критерия (5) выступал минимум взвешенных относительно приоритета отказов в обслуживании на транспортном уровне ТКС. Сама же оптимизация осуществляется с учетом динамических ограничений (1), а также ограничений по временным показателям QoS (6) и (7), по показателям надежности (8) путем численного расчета таких важных параметров как величины MSS, интервала RTT, режимов передачи в TCP, а также параметров алгоритма борьбы с перегрузками.

В ходе постановки оптимизационной задачи, формулировки условий ограничений (1), (6) – (8) произведено удовлетворение таких важных требований к протоколу TCP, как учет динамичности, многопоточности и мультирежимности процесса передачи данных; обеспечение согласованности с работой других средств борьбы с перегрузкой с возможностью выбора той или иной модели блокировки пакетов. Дополнительный учет условий обеспечения устойчивости, полученных в работе [11], позволит снизить чувствительность качества получаемых решений

относительно возможных колебаний исходных данных, например, величины доступной пропускной способности каналов, числа TCP-потоков и др.

Список литературы

1. Бакланов И.Г. NGN: принципы построения и организации / И.Г. Бакланов; под ред. Ю.Н. Чернышова. – М.: Эко-Трендз, 2008. – 400 с.
2. Семенов Ю.В. Проектирование сетей связи следующего поколения / Ю.В. Семенов. – СПб.: Наука и Техника, 2005. – 240 с.
3. Стивенс У.Р. Протоколы TCP/IP. Практическое руководство / У.Р. Стивенс; пер. с англ. и комментарии А.Ю. Глебовского. – СПб.: «Невский диалект»-«БХВ-Петербург», 2003. – 672 с.
4. Куроуз Джеймс Ф. Компьютерные сети: многоуровневая архитектура Интернета / Джеймс Ф. Куроуз, Кит В. Росс. – 2-е изд. – СПб.: Питер, 2004. – 765 с.
5. Математичні основи теорії телекомунікаційних систем / Під заг. ред. В.В. Поповського. – Х.: ТОВ «Компанія СМІТ», 2006. – 564 с.
6. Вегешна Ш. Качество обслуживания в сетях IP: пер. с англ. / Ш. Вегешна. – М.: Изд. дом «Вильямс», 2003. – 368 с.
7. Вишневский В.М. Теоретические основы проектирования компьютерных сетей / В.М. Вишневский. – М.: Техносфера, 2003. – 512 с.
8. Крылов В.В. Теория телетрафика и ее приложения / В.В. Крылов, С.С. Самохвалова. – СПб.: БХВ-Петербург, 2005. – 288 с.
9. Анищенко В.С. Устойчивость, бифуркации, катастрофы / В.С. Анищенко // Соросовский образовательный журнал. – 2000. – № 6. – С. 105-109.
10. Томпсон Дж. М.Т. Неустойчивости и катастрофы в науке и технике: пер. с англ. / Дж. М.Т. Томпсон. – М.: Мир, 1985. – 254 с.
11. Старкова Е.В. Анализ устойчивости TCP-сеанса в рамках нелинейных моделей / Е.В. Старкова, М.В. Меркулов // Перспективные технологии в средствах передачи информации: Материалы 8-й международной научно-технической конференции; Владим. гос. ун-т; редкол.: А.Г. Самойлов (и др.). – Владимир, 2009. – С. 145-149.

Поступила в редколлегию 29.03.2010

Рецензент: д-р техн. наук, проф. В.В. Поповский, Харьковский национальный университет радиоэлектроники, Харьков.

МЕТОД ОПТИМІЗАЦІЇ ПРОЦЕСУ ПЕРЕДАЧІ ДАНИХ У РАМКАХ TCP-СЕАНСІВ

О.В. Лемешко, О.В. Старкова

У роботі проведено аналіз існуючих версій протоколу TCP, а також перспективних рішень в області управління передачею пакетів на транспортному рівні Еталонної моделі взаємодії відкритих систем (EMBBS). Виявлено ряд недоліків математичних моделей і методів, покладених в основу даного протоколу, усунення яких пов'язане з оптимізацією відповідних мережевих параметрів. Запропоновано підхід до рішення задачі оптимізації, заснований на динамічній моделі TCP-сеансів і ряд умов-обмежень, направлених на забезпечення заданих показників якості і стійкості TCP-сеансів, що враховує.

Ключові слова: TCP-сеанс, процес передачі даних.

METHOD OF OPTIMIZATION OF PROCESS OF COMMUNICATION OF DATA WITHIN THE FRAMEWORK OF TCP-SESSIONS

O.V. Lemeshko, O.V. Starkova

The analysis of existent versions of protocol of TCP is conducted in work, and also perspective decisions in area of management the transmission of packages at transport level of the Standard model of co-operation of the open systems. The row of lacks of mathematical models and methods, fixed in basis of this protocol is exposed, the removal of which is related to optimization of the proper parameters of networks. Approach is offered to the decision of task of optimization, based on the dynamic model of TCP-sessions and taking into account row of terms-limitations, directed on providing of the set indexes of quality and stability of TCP-sessions.

Keywords: TCP-session, process of communication of data.