

АДАПТИВНОЕ ОГРАНИЧЕНИЕ ИНТЕНСИВНОСТИ ТРАФИКА НА ПРИГРАНИЧНЫХ УЗЛАХ МУЛЬТИСЕРВИСНОЙ СЕТИ СВЯЗИ

ВВЕДЕНИЕ

Несмотря на стремительное развитие технологий транспортных телекоммуникационных сетей (ТКС) и, прежде всего, технологий волоконно-оптической связи, борьба с перегрузками все еще остается достаточно актуальной сетевой задачей. При высокой загрузке ТКС обеспечить заданные значения показателей качества обслуживания (Quality of Service, QoS) можно лишь за счет ограничения интенсивности трафика, поступающего на приграничные узлы сети [1]. Учитывая мультисервисный характер современных ТКС, ограничение интенсивности или отказы в обслуживании должны касаться, как правило, лишь низкоприоритетных трафиков на перегруженных участках сети. Таким образом, процесс ограничения интенсивности трафика должен носить адаптивный характер, основываясь на постоянном мониторинге состояния ТКС – загрузки трактов передачи и сетевых узлов, характеристик трафиков пользователей и требований к качеству их обслуживания. Немаловажным требованием также является обеспечение максимального уровня согласованности работы средств ограничения трафика с другими средствами борьбы с перегрузками, к числу которых можно отнести протоколы маршрутизации, механизмы ограничения длин очередей и др.

АНАЛИЗ ИЗВЕСТНЫХ СРЕДСТВ ОГРАНИЧЕНИЯ НАГРУЗКИ НА ПРИГРАНИЧНЫХ УЗЛАХ

В современных мультисервисных ТКС задачи ограничения интенсивности трафика решаются с помощью механизмов Traffic Shaping (TS) и Committed Access Rate (CAR). В основу этих средств положены алгоритмы Leaky Bucket ("дырявое ведро") и Token Bucket ("маркерное ведро"). Несмотря на сходное предназначение, механизмы CAR и TS отличаются способом обработки трафика в момент исчерпания маркеров. Traffic Shaping представляет собой механизм сглаживания поступающего на сетевой узел пользовательского трафика с целью недопущения перегрузки исходящих трактов передачи и удовлетворения QoS-требований. В соответствии с механизмом TS интенсивность пульсирующего трафика выравнивается до согласованной скорости передачи информации (CIR) путем постановки в очередь (буферизации) пакетов, интенсивность передачи которых превысила среднее значение. Буферизованные пакеты передаются по мере накопления достаточного числа маркеров. Передача поставленных в очередь пакетов планируется механизмом обслуживания очередей, например, взвешенным механизмом справедливого обслуживания очередей (Weighted Fair Queuing, WFQ). Трафик, интенсивность которого превышает CIR, может быть отклонен в случае переполнения сети. Реализуя функцию ограничения трафика, механизм CAR не помещает пакеты в буфер и не сглаживает трафик, что может привести к отбрасыванию пакетов в моменты превышения максимально допустимого размера всплеска [2, 3].

Таким образом, существующие средства ограничения интенсивности поступающего в сеть трафика носят распределенный характер, основываясь на информации о средней скорости поступления пакетов в сеть, заявленной в договоре о QoS. Кроме того, эвристические по своему содержанию алгоритмы корзины маркеров и дырявого ведра не способны учесть изменение текущей загрузки сетевого узла и характеристик трафиков прочих пользователей; а также не согласованы с работой других средств управления трафиком. Это значительно снижает функциональность механизмов TS и CAR и существенно ограничивает область их применения в мультисервисных ТКС. В этой связи, актуальной научной и практической задачей является разработка моделей и методов адаптивного ограничения интенсивности трафика, поступающего на приграничные узлы ТКС, которые могли бы быть положены в основу соответствующих технологических решений – протоколов и механизмов управления трафиком.

МОДЕЛИ АДАПТИВНОГО ОГРАНИЧЕНИЯ ИНТЕНСИВНОСТИ ТРАФИКА НА ПРИГРАНИЧНЫХ УЗЛАХ МУЛЬТИСЕРВИСНОЙ СЕТИ СВЯЗИ

Максимального уровня согласованности решений задач управления сетевыми ресурсами и, в том числе, информационным трафиком можно обеспечить лишь основываясь на единой (комплексной) математической модели управления. За основу подобной комплексной модели целесообразно выбрать математическую модель маршрутизации, обладающей по сравнению с моделями других процессов управления наиболее высокой степенью системности, т.е. учитывающей большее число структурных и функциональных сетевых параметров. Основываясь на модели маршрутизации, необходимо максимально адекватно и математически корректно провести формализацию остальных процессов управления – распределения канальных и буферных ресурсов, ограничения интенсивности трафика, обслуживания очередей и др. К числу наиболее результативных попыток подобного развития модели маршрутизации является динамическая модель управления маршрутизацией, канальным и буферным ресурсом [4], а также модель комплексного решения задач маршрутизации, распределения канального ресурса с целью обеспечения заданных показателей качества обслуживания [5].

Важным моментом при выборе математической модели маршрутизации, по аналогии с [4] и [5], является то, что она должна носить потоковый характер, в отличие от графовых моделей, положенных в основу большинства современных протоколов однопутевой маршрутизации [6]. Несмотря на достаточно большое число вариантов потокового описания процессов маршрутизации, основные отличия в целом касаются формы представления условия сохранения потока в сетевых узлах и ТКС в целом. Ниже будут предложены дополнения к наиболее известным потоковым версиям математических моделей маршрутизации с целью формализации процессов ограничения интенсивности трафика, поступающего в сеть.

Дополнение математической модели маршрутизации (пример 1)

Пусть структура сети описывается с помощью графа $G = (M, E)$, где M – множество узлов, а E – множество трактов передачи. Для каждой дуги $(i, j) \in E$ характерна ее пропускная способность φ_{ij} . Пусть также K – множество трафиков, которые циркулируют в ТКС. Тогда для $k \in K$ необходимо указать r_k , s_k и d_k – интенсивность k -го трафика, узел-источник и узел-получатель соответственно. Тогда для каждого тракта $(i, j) \in E$ и обслуживаемого трафика $k \in K$ величина x_{ij}^k характеризует долю k -го трафика, протекающего в тракте (i, j) . В дополнение к моделям маршрутизации [7, 8] вводится также величина α^k , которая представляет собой долю k -го трафика, получившего отказ в обслуживании сетью. Тогда вектор искомых параметров можно представить в следующей форме:

$$X = \begin{bmatrix} x_{ij}^k \\ - \\ \alpha^k \end{bmatrix}, (i, j) \in E, k \in K. \quad (1)$$

Размерность вектора X определяется произведением количества в сети трактов передачи и числа анализируемых трафиков пользователей. В ходе расчета вектора (1) необходимо минимизировать функцию

$$\min_X c^t X, \quad (2)$$

которая численно характеризует затраты на управление сетевыми ресурсами, а координаты вектора

$$c = \begin{bmatrix} c_{ij}^k \\ \dots \\ c^k \end{bmatrix}, \quad (3)$$

в свою очередь, определяют величину удельного штрафа за загруженность трактов передачи ТКС (c_{ij}^k) и за ограничение в обслуживании трафиков пользователей (c^k). Для трафиков с более высоким приоритетом должны быть установлены и более высокие штрафы за отказ, что позволит обеспечить адаптивный характер процессу отказов в обслуживании. Размерность вектора (3) полностью совпадает с размерностью искомого вектора X (1).

Для корректного решения маршрутных задач и задачи ограничения интенсивности трафика в рамках предлагаемой комплексной модели необходимо обеспечить выполнение условий сохранения потока в сетевых узлах и сети в целом:

$$\begin{cases} \sum_{j:(i,j) \in E} x_{ij}^k - \sum_{j:(j,i) \in E} x_{ji}^k = 0 & \text{при } k \in K, i \neq s_k, d_k; \\ \sum_{j:(i,j) \in E} x_{ij}^k - \sum_{j:(j,i) \in E} x_{ji}^k = 1 - \alpha^k & \text{при } k \in K, i = s_k; \\ \sum_{j:(i,j) \in E} x_{ij}^k - \sum_{j:(j,i) \in E} x_{ji}^k = \alpha^k - 1 & \text{при } k \in K, i = d_k, \end{cases} \quad (4)$$

а также условий предотвращения перегрузки трактов передачи ТКС:

$$\sum_{k \in K} r_k x_{ij}^k \leq \varphi_{ij}; (i, j) \in E. \quad (5)$$

В общем случае, в соответствии с физикой решаемой задачи (1)-(5) на координаты α^k вектора X необходимо наложить следующие ограничения:

$$0 \leq \alpha^k \leq 1, \quad (6)$$

если в соответствии с договором о качестве обслуживания (Service Level Agreement, SLA) допускается частичное ограничение скорости доступа, или

$$\alpha^k \in 0,1 \quad (7)$$

– в противном случае.

Дополнение математической модели маршрутизации (пример 2)

Пусть r_{ij} – интенсивность входного трафика, поступающего в сеть через i -й узел и адресованный j -му узлу, тогда дополнение математической модели маршрутизации, предложенной Галлагером [9, 10], состоит во введении в уравнение сохранения потока новой управляющей переменной α_{ij} , численно характеризующей долю интенсивности трафика r_{ij} , получившей отказ в обслуживании:

$$\gamma_{ij} = r_{ij}(1 - \alpha_{ij}) + \sum_{s \in M_i} \gamma_{sj} \phi_{ji}^s \quad (8)$$

$$\text{при } \sum_{j \in M_i} \gamma_{ij} \phi_{js}^i = \lambda_{ij} \quad (i, j = \overline{1, m}, i \neq j), \quad \phi_{js}^i = \begin{cases} 0, & \text{если } i = j; \\ \geq 0, & \text{если } i \neq j \end{cases} \quad \text{и} \quad \sum_{s \in M_i} \phi_{js}^i = 1,$$

где M_i – множество соседних узлов i -му узлу; γ_{ij} – интенсивность потока в i -м узле, определяемая как сумма входного потока и потока, поступающего на i -й узел от смежных узлов для j -го узла (1/с); ϕ_{js}^i – маршрутная переменная, численно характеризующая долю потока γ_{ij} , протекающего из i -го узла по тракту (i, s) ; λ_{ij} – интенсивность трафика в тракте (i, j) (1/с).

Согласно физическому смыслу вновь вводимых управляющих переменных, на них необходимо наложить ограничения подобные (6) и (7): $0 \leq \alpha \leq 1$ либо $\alpha \in \{0, 1\}$.

Комплексный характер расчета искомым переменных γ_{ij} , ϕ_{js}^i , α_{ij} , регламентирующих порядок решения задач маршрутизации и ограничения интенсивности трафика, влечет за собой и пересмотр критериев оптимальности реализуемого управления, т.к. традиционная минимизация среднесетевой задержки или максимальной загруженности трактов передачи ТКС [9, 10] непременно повлечет в рамках модели (8) полное ограничение интенсивности трафика ($\alpha \rightarrow 1$). Разумный компромисс может быть достигнут лишь благодаря введению переменных α_{ij} в целевую функцию, по аналогии с моделью (1)-(3).

В соответствии с предложенным подходом введение переменных α_{ij} в модель [5] позволит получить согласованное решение задач маршрутизации, распределения канального ресурса и ограничения интенсивности поступающего в сеть трафика с обеспечением заданных показателей качества обслуживания.

ПРИМЕР РЕШЕНИЯ КОМПЛЕКСНОЙ ЗАДАЧИ МАРШРУТИЗАЦИИ И АДАПТИВНОГО ОГРАНИЧЕНИЯ ТРАФИКА

Решение комплексной задачи маршрутизации и адаптивного ограничения интенсивности трафика, поступающего в сеть, продемонстрируем для большей наглядности для однопродуктового случая, т.е. для случая обслуживания одного трафика, что не снижает общности полученных результатов. Пусть структура ТКС, состоящей из пяти узлов и шести трактов передачи (с указанием условной величины их пропускной способности), представлена на рис.1. Первый узел – узел-отправитель, а пятый – узел-получатель.

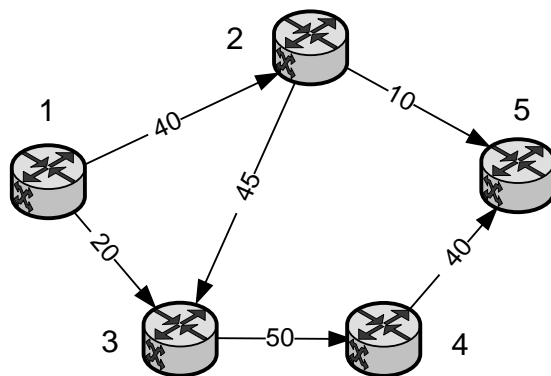


Рис.1

Тогда для случая $c = 10 \ 10 \ 10 \ 10 \ 10 \ 100 \text{ }^{-t}$ (3) получены следующие результаты решения при отсутствии ограничения трафика (рис.2 а), а также в случае отказов в обслуживании (рис.2 б) с указанием в трактах передачи интенсивности передаваемого трафика.

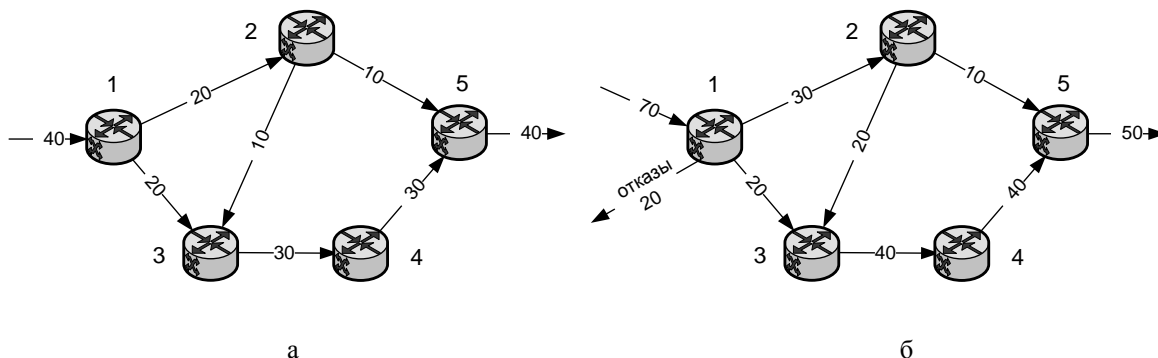


Рис.2

Из полученных результатов расчетов видно, что при отношении весовых коэффициентов c_{ij}^k и c^k как 1:10, ограничение интенсивности трафика наблюдается лишь в условиях перегрузки сети, когда интенсивность поступающего трафика превосходит пропускную способность сети, которая для приведенных исходных данных составляет 50 1/с. При уменьшении соотношения c_{ij}^k и c^k до 1:4 (рис.3 а) или 1:3 (рис.3 б) наблюдается превентивное ограничение нагрузки при свободных канальных ресурсах.

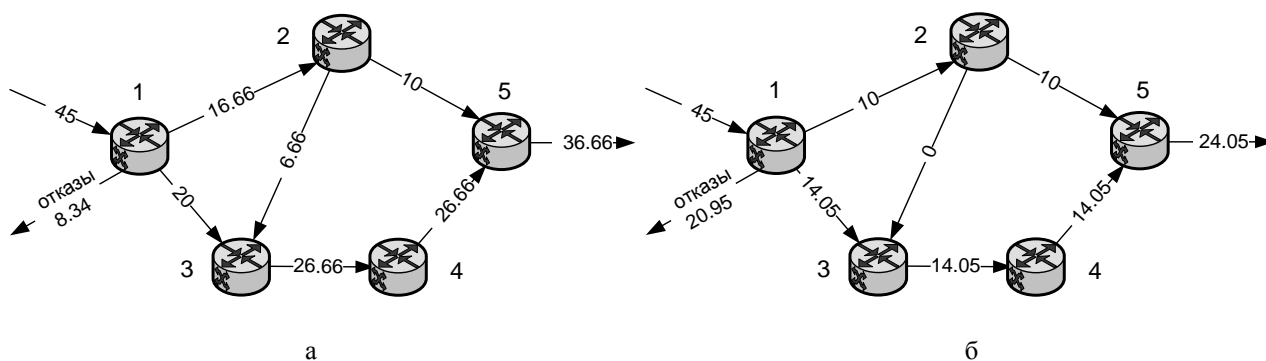


Рис.3

Обоснованный выбор соотношения весовых коэффициентов c_{ij}^k и c^k , отвечающих соответственно за расчет маршрутных переменных и переменных α^k , представляет собой самостоятельную достаточно сложную задачу, в ходе которой необходимо учесть структуру сети (количество узлов и трактов передачи), число обслуживаемых трафиков и их приоритет, а также требования к качеству обслуживания каждого из трафиков. Для упрощения изложения материала на рассмотренных примерах (рис.2 и рис.3) принимались во внимание только скоростные показатели качества обслуживания – средняя скорость (интенсивность) поступления пакетов в сеть.

ВЫВОДЫ

Таким образом, в работе предложен подход к решению задачи адаптивного ограничения интенсивности поступающего в сеть трафика. Преимущество предложенного подхода заключается в обеспечении согласованности решений задач ограничения интенсивности тра-

фика с другими задачами управления (маршрутизации, распределения канального ресурса) и обеспечения качества обслуживания. Необходимый уровень согласованности решений основных задач сетевого управления был достигнут на основе усовершенствования (дополнения) математических моделей маршрутизации. Усовершенствование заключалось во введении в уравнения (условия) сохранения потока дополнительного множества управляющих переменных, отвечающих за адаптивное ограничение интенсивности поступающего в сеть трафика. Комплексный характер модели основывался на одновременном (согласованном) расчете управляющих переменных, отвечающих как за маршрутизацию, так и за задачи адаптивного ограничения трафика.

Экспериментальным путем установлено, что отношение коэффициентов c_{ij}^k и c^k , отвечающих за загруженность сети и отказы в обслуживании, существенным образом влияет на характер возможных отказов в обслуживании. Чем это соотношение больше, тем более запоздалой является реакция сети на возможную перегрузку; чем соотношение ниже, тем выше вероятность необоснованного отказа в обслуживании поступающего в сеть трафика.

В этой связи, развитие предложенного подхода видится в разработке методики рационального (оптимального) выбора значений весовых коэффициентов (3) с обоснованием требуемой величины их соотношения с целью обеспечения превентивного характера процессов ограничения интенсивности поступающего в сеть трафика в порядке роста его важности (приоритетности). Область преимущественного использования предложенных моделей – мультисервисные сети связи с поддержкой качества обслуживания.

Список литературы: 1. *Гургенидзе А.Т., Кореш В.И.* Мультисервисные сети и услуги широкополосного доступа. С-Пб.: Наука и Техника. 2003. 400 с. 2. *Справочник по телекоммуникационным технологиям: Пер. с англ.* М.: Издательский дом «Вильямс», 2004. 640 с. 3. *Вегенша Ш.* Качество обслуживания в сетях IP: Пер. с англ. М.: Издательский дом «Вильямс», 2003. 386 с. 4. *Лемешко А.В., Беленков А.Г.* Динамическая модель комплексного решения задач маршрутизации и абонентского доступа в территориально-распределенных телекоммуникационных сетях // Открытые информационные и компьютерные интегрированные технологии. Сб. науч. трудов Нац. аэрокосмич. ун-та им. Н.Е. Жуковского „ХАИ“. Вып. 18, Харьков, НАКУ, 2003. С. 134-139. 5. *Дробот О.А.* Комплексная модель обеспечения гарантированного качества обслуживания с реализацией динамических стратегий распределения сетевых ресурсов // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2007. № 148. С.43-54. 6. *Остерлох Х.* Маршрутизация в IP-сетях. Принципы, протоколы, настройка. С.Пб.: BHV.-С.Пб., 2002. 512 с. 7. *Seok Yo., Lee Yo., Choi Ya., Kim C.* A constrained multipath traffic engineering scheme for MPLS networks // Proc. of IEEE ICC 2002. New York. 2002. P. 2431-2436. 8. *Wang Y., Wang Z.* Explicit routing algorithms for Internet Traffic Engineering // Proc. of 8th International Conference on Computer Communications and Networks. Paris, 1999. P. 582-588. 9. *Бертсекас Д., Галлагер Р.* Сети передачи данных. М.: Мир, 1989. 544 с. 10. *Gallager R. G.* A minimum delay routing algorithm using distributed computation // IEEE Trans. on communications. 1975. Vol. 25, №1. P. 73-85.

*Харьковский университет Воздушных Сил
Имени Ивана Кожедуба*

Поступила в редколлегию