DOI: 10.17117/na.2016.06.02.110

<http://ucom.ru/doc/na.2016.06.02.110.pdf>

Поступила (Received): 04.06.2016

Москалец Н.В., Поповская Е.О., Тарасов К.А.
Оптимизация суммарного взвешенного времени
обслуживания в пиринговой сети

Moskalets N.V., Popovskaya E.A., Tarasov K.A.
Optimization of total weighted service time in the P2P network

Описанные модели и методы теории расписаний позволяют адекватно отображать структуру и функцию пиринговой сети и учитывать их динамику, ограничение предшествований, перестановки, прерывания и другие типовые технологические решения, что может быть использовано для повышения производительности сети и качества предоставления услуг QoS, QoE

Ключевые слова: пиринговая сеть, теория расписаний, качество обслуживания

Described models and methods of scheduling theory can adequately display the structure and function of peer to peer networks and take into account their dynamics, precedence constraints, permutations, interrupts, and other typical technological solutions that can be used to improve network performance and quality of service QoS, QoE

Key words: P2P-network, scheduling theory, quality of service

Москалец Николай Вадимович

Кандидат технических наук, доцент
Харьковский национальный университет
радиоэлектроники
Украина, г. Харьков, проспект Науки, 14

Moskalets Mykola Vadymovych

Candidate of Engineering Sciences, Associate
Professor
Kharkiv national university of radio electronics
Ukraine, Kharkiv, Nauky ave., 14

Постановка проблемы: На основе разработки математической модели в сетях пиринговой связи дается формулировка задачи оптимизации суммарного взвешенного времени обслуживания.

Методы: Оптимизация математической модели, построенной на теории расписаний.

Результаты: Получено общесистемное решение, позволяющее оптимизировать время обслуживания пиринговой сети.

Практическая значимость: Описанные модели и методы теории расписаний позволяют адекватно отображать структуру и функцию пиринговой сети и учитывать их динамику, ограничение предшествований, перестановки, прерывания и другие типовые технологические решения, что может быть использовано для повышения производительности сети и качества предоставления услуг QoS, QoE.

Введение

Приложения Peer-To-Peer (P2P) стали очень популярны среди интернет-пользователей. P2P технологии улучшают масштабируемость системы при низ-

ких затратах реализации. В отличие от традиционных клиент-серверных архитектур, пиры в сети выступают в качестве клиента (лич-скачивающий) и сервера (сид-раздающий). Пир не только скачивает файл из сети, но и раздает загруженный файл другим пользователям в сети. Части файлов обмениваются через прямые соединения между пирами.

Другим важным свойством пиринговых сетей является возможность упорядоченности фрагментов, прерываний, вхождений и выходов из сети в любой момент времени.

Практическое применение пиринговых сетей постоянно расширяется, а пользовательский трафик по данным Cisco Systems в 2020 году достигнет 60 петабайт/месяц [1]. Большую практическую пользу дает возможность организации распределенных вычислений, что позволяет этим сетям конкурировать с производительностью современных суперкомпьютеров. Значительная доля задач приходится на организацию он-лайн игр (MMOG). Однако, по тем же данным [1], около половины трафика пиринговых сетей приходится на организацию файлообмена P2P-TV, что привлекает к этой технологии особое внимание.

Архитектура пиринговых сетей может быть структурированной или неструктурированной [2]. В структурированных системах P2P-HQTV структура оверлея обычно представляет полудетерминированную топологию, что вызывает сложности в реструктуризации при присоединении или отсоединении очередного пира. Отличительной особенностью неструктурированной архитектуры P2P-TV систем является то, что топология оверлея полностью отделена от топологии распределения. В результате образуется mesh-топология позволяющая реализовать высокую степень динамичности и децентрализации.

Пиринговые сети, являясь частными, оверлейными файлообменными сетями, обладают рядом специфических особенностей по отношению к общей IP-сети. Так в известных файлообменных приложениях (Bit Torrent) оптимизация производительности сети сводится к поиску «лучшего» пира из которого загружается содержимое. Укажем на другие важные для повышения производительности сети задачи, которые необходимо решать параллельно в реальном времени [3, с.150]: алгоритмы планирования элементов, выбор пира с низкими задержками, обеспечение взаимодействия между динамикой оверлея и IP-маршрутизацией.

Таким образом, общесистемные критерии качества: QoS и QoE, хотя и имеют непосредственную зависимость от качества решения перечисленных и многих других задач, не всегда позволяют получить решение в целом. Поэтому многие работы, посвященные исследованиям пиринговых сетей носят частный характер [4, с.10]. Наша задача состоит в попытке получения общесистемного решения, оптимального по критерию минимума задержки при завершении обслуживания, при наличии ряда типовых ограничений [5-7].

Постановка задачи:

Будем считать, что P2P-TV сеть обеспечивает одновременный просмотр n m – программ $m \in M$. Одну из этих программ просматривает l -абонентов (пи-

ров) $l \in L$. В любой момент каждый из l -пиров имеет право прекратить обслуживание, а затем – также начать. У каждого пира имеется $N = \{1, 2, \dots, n\}$ требований, представимых в виде отдельных j -фрагментов скачиваемого контента.

Длительность обслуживания сетью каждого из перечисленных j -фрагментов составляет $p_j p_j'$, где $j \in N$. Процесс скачивания контента представляет собой последовательность процедур с обеспечением частичной упорядоченности, что в конечном счете определяется ограничениями предшествования. В случае если в ограничениях имеется директивный срок завершения обслуживания по j -фрагменту d_j , то разница между моментом завершения C_j и директивным моментом d_j при $C_j > d_j$ считается запаздыванием в обслуживании или обслуживанием с опережением при $C_j < d_j$.

Будем считать, что каждый пользователь обладает буфером, способным содержать все множество требований $N = \{1, 2, \dots, n\}$, где n - количество мест для хранения n -фрагментов, полученных на предыдущих шагах, плюс одно место, предназначенное для получения очередной новой порции данных. В соответствии с правилом загрузки, после того, как сервер случайным образом выбрал пира для начальной загрузки, он находит в сети другого пира и пытается загрузить от него недостающую порцию данных. Таким образом, в результате различий в пропускных способностях каналов и длительностях загрузки и приема буфера могут иметь место случаи нарушения частичного порядка [8, с15-27].

Для учета ограничений предшествования в обслуживании зададим бинарное отношение строгого порядка на множестве требований N . Требование i является предшественником требования j , а требование j - последователем потомком по отношению к i . Данное отношение задается ориентированным бесконтурным графом G , множество вершин (узлов) которого идентифицируется с множеством требований, а путь из вершины i в вершину j существует, когда $i \rightarrow j$.

Формулировка задачи минимизации суммарного взвешенного времени обслуживания.

Будем считать, что при реализации N -го требования на терминал поступает не более одного фрагмента одновременно. Для каждого j -го фрагмента, принадлежащего множеству $N = \{1, 2, \dots, n\}$ заданы параметры:

- длительность обслуживания j -фрагмента – p_j , $0 < p_j^L \leq p_j^U$; где индексы L и U относятся соответственно к нижней и верхней границам обслуживания. Обычно длительность обслуживания p_j оказывается неизвестной вплоть до момента завершения данного j -обслуживания. Случайный характер p_j определяется различием в параметрах линий скачивания и используемых технологий;

- каждому j -требованию сопоставляется весовой множитель $w_j > 0$. Значения весовых коэффициентов w_j не могут быть одинаковыми, в первую очередь из-за того, что важность последующих фрагментов к концу скачивания возрастает, поэтому $w_n \geq w_{n-1} \geq \dots w_1$.

Для нахождения решений задачи воспользуемся результатами теории расписаний (Scheduling Theory) [9, с.80; 10, с.21], которая в частности сводится к построению функции, определенной на множестве перестановок и к разработке оптимизационных алгоритмов. На каждом из n – этапов обслуживания возможны различные длительности реализации p_j каждого из j – обслуживаний, $j \in N$. На практике значение p_j образуют счетное множество вариантов из интервала $[p_j^L; p_j^U]$. Таким образом можно рассматривать k – мерный вектор

$$p_j^T = (p_1, p_2, \dots, p_k), p_j \in [p_j^L; p_j^U]. \quad (1)$$

Значения p_j в общем случае становятся известны после приема каждого из фрагментов: $p_j = p_{ps}^{(j)}$.

Пусть P – обозначает множество всех векторов (1) возможных длительностей обслуживания требований

$$p_i^T = (p_1^{(l)}, p_2^{(k)}, \dots, p_n^{(k)}), k = 1, 2, \dots, k. \quad (2)$$

Вектор (2) представляет собой сценарий событий, последовательности n – фрагментов по заявке данного пира.

Очевидно P – замкнутый прямоугольник неотрицательных чисел.

При планировании расписания S следования фрагментов скачиваемого контента и реализации сценария (2) в качестве критериальной можно выбрать аддитивную взвешенную линейную функцию:

1. Суммарное взвешенное время, затрачиваемое для приема вектора (2)

$$f_i(C_i) = \sum_{i=1}^n w_i C_i, \quad (3)$$

где w_i – весовой множитель, определяющий значимость i – го фрагмента, в том числе времени его обслуживания.

Для различных w_i значения (3) функций $f_i(C_i)$ представляется в виде прямых на графиках рис.1а.

2. Суммарное взвешенное количество фрагментов, принятых с опозданием $f_i(C_i) = \sum_{i=1}^n w_i U_i$,

где U_i равна нулю, если $C_i \leq d_i$ – директивный срок (в отсутствии опозданий) и равна единице при $d_i \leq C_i$.

Возможные значения функций $f_i(C_i)$ представлены на рис.1б.

3. Суммарное взвешенное опоздание фрагментов

$$f_i(C_i) = \sum_{i=1}^n w_i T_i, \quad (4)$$

где $T_i = \max(0, C_i - d_i)$.

Значения функции (5) представлены на рис. 1 в. а-(3); б-(4); в-(5).

Следует иметь в виду, что сценарий (2) в общем случае остается неопределенным вплоть до момента завершения обслуживания по всем i и j . При этом из-за NP – полноты минимизация критериальных функций (3),(4),(5) невозможна по крайней мере на стадии вычисления и составления его оптимального расписания.

Рис. 1. График качественного поведения критериальных функций

Вместе с тем, для основных технологий P2P-TV, при скачивании конкретных файлов, дифференциальная фрагментация как правило не используется. Это позволяет выбирать для всех p_i

$$p_i = p, \forall i. \quad (5)$$

С условием (5) можно считать параметры вектора (1) известным априори. При этом удастся уйти от NP -полноты и решить оптимизационную задачу за время $O(n \log n)$ [10;11, с.577]:

$$f_{opt}(t) = \min_s \sum_{i=1}^n w_i C_i(p_i = p). \quad (6)$$

В теории расписаний имеется множество расписаний подобных задач при условии дополнительных ограничений. Так известны решения, основанные на методе динамического программирования [12]. Среди других практически важных следует указать на методы дихотомии [12,13], робастные методы. Перспективным является метод, основанный на анализе чувствительности [8,с.20], предлагающий двухстадийное принятие решений: стадию априорного планирования k -вариантов в режиме off-line и стадию составления расписаний в режиме on-line. В силу различий условий скачивания контента, других случайных факторов в последовательности принимаемых пиром фрагментов могут быть нарушены условия предшествования, что потребует упорядочения и соответствующих перестановок в порядке следования фрагментов. Это накладывает дополнительные ограничения на решение оптимизационной задачи. Так выбранный нами ранее критерий (3) с учетом необходимой процедуры перестановки $\pi = (\pi_1, \pi_2, \dots, \pi_{n!})$ приобретает вид:

$$J = \sum_{j=1}^n w_j C_j(\pi, P) = \min_{\pi_k} \left\{ \sum_{\pi} \sum_j w_j C_j(\pi_k p) \right\} \quad (7)$$

Очередным важным ограничением, встречающемся в процессе функционирования пиринговых сетей, является наличие прерываний, вызванных технологическими причинами при обслуживании в условиях большой нагрузки или за счет влияний более приоритетных обслуживаний [8].

Рассмотрим пример оптимизации среднего взвешенного времени начала обслуживания. Будем считать M – длительность обслуживания скачиваемого файла известной. Исходные данные:

$I=\{i\}$ – множество фрагментов; $P=\{p_i\}$ – множество длительностей обслуживания; $W=\{w_i\}$ – множество штрафов требований.

Целевая функция: $F(X)=W^T * X \rightarrow \min$

Для каждой пары фрагментов $i, j \in I$ выполняется одно из условий $x_i - x_j \geq p_j$, что соответствует предшествованию j по отношению к фрагменту i или $x_j - x_i \geq p_i$, когда i предшествует выполнению j . Пусть также коэффициент предшествования равен $y_{i,j}=1$, если i предшествует j и 0 в противном случае. Это позволяет записать систему ограничений в виде:

$$(M + p_j)y_{ij} + (x_i - x_j) \geq p_j, \quad (8),$$

$$(M + p_i)(1 - y_{ij}) + (x_j - x_i) \geq p_i. \quad (9)$$

Таким образом, получена следующая целевая функция:

$$F(X)=4x_1 + 2x_2 + 2x_3 + 3x_4 \rightarrow \min.$$

С использованием пакета Matlab получено оптимальное решение для четырех фрагментов при предположении: $p=\{6,2,2,3\}$; $W=\{4,2,2,3\}$.

Полученное оптимальное решение, представленное на рис.2 диаграммой Гантта в виде последовательности.

Рис. 2. Оптимальная последовательность скачиваемых фрагментов

Заключение

1. Пиринговые P2P-TV сети интенсивно развиваются и приобретают все большую популярность как в области индустрии развлечений, так и при решении ряда профессиональных задач инфокоммуникационного направления.

2. Существующие технологии пиринговых P2P-TV сетей позволяют находить решения по оптимизации имеющихся ресурсов, направленных на повышение производительности сети и качества предоставления услуг QoS, QoE.

3. Модели и методы теории расписаний являются адекватными при решении задач в области улучшения качественных показателей пиринговых сетей. Они позволяют адекватно отображать как структуру, так и функции этих сетей, учитывать их динамику, ограничение предшествований, перестановки, прерывания и другие типовые технологические условия.

Список используемых источников:

1. Cisco Visual Networking Index; Forecast and Methodology // Cisco Public, 2012-2017.
2. Tarkoma S. Overlay Network: Toward Information Networking // CRC Press, 2010.
3. Setton E., Girod B. Peer-to-Peer Video Streaming // Springer. 2007. 150 p.
4. Адами Амину, Гайдамака Ю.В. Аппроксимация вероятностных характеристик модели сети P2P // Л.2013. Инф.упр.сети. С.4-10.
5. Wehrle K., Gunes M., Gross J. Modeling and tools for Network simulation // Springer, Heidelberg, Dordrecht, London, New York. 2010.

6. Дорт-Гольц А.А. Анализ функционирования наложенных сетей в сетях операторов // М.Электросвязь №3, 2014. Стр.22-27.
7. Moskalets, N.V., Vasylenko, Yu.A. Analysis of probability characteristics of the combined multiple access // CriMiCo 2013 – 2013 23rd International Crimean Conference Microwave and Telecommunication Technology, Conference Proceedings, pp.466.
8. Гордон В.С., Долгий А.Б. Задачи теории расписаний с заданным частичным порядком // М. Автоматика и телемеханика №10, 2010. Стр.15-27.
9. Лазарев А.А., Кварцхелия А.Г. Свойство оптимальных расписаний задачи теории расписаний минимизации суммарного взвешенного момента окончания для одного прибора // М. Автоматика и телемеханика №10, 2010. Стр.80-90.
10. Baptiste Ph. Scheduling Equal-Length Jobs on Identical Parallel Machines // Discrete Appl. Math. 2000, №103, p. 21-32.
11. Engels D.W., Feldman Y., Karger D.R. Parallel processor scheduling with delay constraints // SODA, 2001, p. 577-585.
12. Pinedo M. Scheduling Theory, Algorithms and Systems // Springer-Verlag, 2008.
13. Brucher P., Knust S. Complex Sheduling // Springer-Verlag, 2006.

© 2016, Москалец Н.В., Поповская Е.О., Тарасов К.А.
 Оптимизация суммарного взвешенного времени
 обслуживания в пиринговой сети

© 2016, Moskalets N.V., Popovskaya E.A.,
 Tarasov K.A.
 Optimization of total weighted service time in the
 P2P network