Стохастическая модель протокола BitTorrent

Парадигмы централизованных и пиринговых коммуникаций исторически сменяют друг друга по мере развития компьютерного оборудования и телекоммуникационных сетей. Сейчас популярны пиринговые методы передачи, позволяющие перемещать большие объемы данных не расширяя возможности существующих линий пермшефдашкуедачи данных. Одним из наиболее распространенных протоколов пиринговых сетей является протокол типа BitTorrent. Основная цель данного протокола — раздать файл как можно большему количеству участников таким образом, чтобы во время распространения доступность файла не уменьшалась. Предложена математическая модель распространения данных по протоколу типа BitTorrent, построенная с помощью аппарата стохастических дифференциальных уравнений. Модель описывает процесс распространения на большое число компьютеров (рабочих станций пользователей) дисплейного класса заранее подготовленного файла образа операционной системы и программного обеспечения. Модель позволяет проанализировать основные характеристики протокола — среднее число личеров, сидеров, среднюю пропускную способность системы и времени загрузки данных. Последняя характеристика особенно важна для анализа одного из подходов к решению одной из основных задач системного администрирования - автоматизации процесса установки образа операционной системы и программного обеспечения на большое число рабочих мест пользователя за ограниченное время - а именно преимущество протокола BitTorrent перед другими возможными решениями задачи.

Ключевые слова: *протокол BitTorrent, пиринговая сеть, математическая модель, система стохастических дифференциальных уравнений, среднее время загрузки данных.*

Кулябов Д.С.,
к.ф.-м.н., доцент кафедры систем телекоммуникаций РУДН, dskulyabov@gmail.com

Королькова А.В.,
к.ф.-м.н., доцент кафедры систем телекоммуникаций РУДН, akorolkova@sci.pfu.edu.ru

Зарядов И.С.,
к.ф.-м.н., старший преподаватель кафедры теории вероятностей и математической статистики РУДН, izaryadov@sci.pfu.edu.ru

1. Введение

Пиринговые методы передачи, позволяющие перемещать большие объемы данных не расширяя возможности существующих линий передачи данных, представляют большой интерес для исследований [1-9]. Одним из наиболее распространенных протоколов пиринговых сетей является протокол BitTorrent, позволяющий раздать файл как можно большему количеству участников таким образом, чтобы во время распространения доступность файла не уменьшалась.

Данная статья посвящена построению с помощью стохастических дифференциальных уравнений [3,10] математической модели распространения данных по протоколу типа BitTorrent. Рассматривается процесс распространения на большое число компьютеров (рабочих станций пользователей) дисплейного класса заранее подготовленного файла образа операционной системы и программного обеспечения. Основная задача строящейся математической модели – с помощью аппарата стохастических дифференциальных уравнений с пуассоновским процессом получить по возможности аналитические выражения основных характеристик протокола – среднее число личеров, сидеров, среднюю пропускную способность системы и среднее время загрузки данных.

Структурно работа состоит из следующих разделов: раздел 2 посвящен постановке задачи, вводятся все основные используемые в работе обозначения, раздел 3 – сама модель, промежуточные и окончательные результаты, раздел 4 – заключение, список литературы.

2. Постановка задачи

Будем рассматривать в системе два типа пиров (peers) – личеры (leechers) (пиры с неполным файлом) и сидеры (seeders) (пиры с полным файлом). Пусть c_i и c_j – число частей передаваемого файла, находящееся у i-го и j-го пиров в момент времени t, всего в системе в рассматриваемый момент времени t находится случайное число $N(t)$ пиров. Величины c_i и c_j принимают значения из множества $\{0, 1, 2, ..., M\}$, где M – общее число частей передаваемого файла.

Будем предполагать, что $N(t)$ – пуассоновский процесс, описывающий число пиров в сети в момент времени t.

Обозначим через P_{ij} вероятность того, что i-й пир получит хотя бы один недостающий кусок файла от j-го пира. Вероятность P_{ij} имеет вид

$$P_{ij} = \begin{cases} 1, & c_i < c_j, \\ 1 - \dfrac{C_{c_i}^{c_j}}{C_M^{c_j}}, & c_i \ge c_j. \end{cases} \quad (1)$$

Стоит отметить, что для вероятности P_{ij} верно следующее неравенство:

$$P_{ij} \ge 1 - \left(\frac{c_i}{M} \right)^{c_j}.$$

Работа выполнена при частичной финансовой поддержке Министерства образования и науки Российской Федерации (проекты 8.7962.2013, 14.U02.21.1874).

Введем следующие обозначения: $X(t)$ – число личеров в системе в момент времени *t*; $Y(t)$ – число сидеров в системе в момент времени *t*; $D_i(t)$ – интенсивность загрузки частей файла *i*-м пиром в момент времени *t*; $U_i(t)$ – интенсивность отдачи частей файла *i*-м пиром в момент времени *t*; μ – средняя интенсивность обмена данными для установленного соединения между двумя пирами (одинакова для всех возможных соединений); ρ – вероятность установления и поддержки соединения между двумя пирами (предполагается одинаковой для всех возможных соединений); B – максимальная пропускная способность установленного соединения (одинакова для всех возможных соединений).

Будем рассматривать два случая:

- нет ограничений на пропускную способность канала,
- наличие ограничений на максимальную пропускную способность канала.

3. Построение математической модели распространения данных по протоколу типа BitTorrent с помощью аппарата стохастических дифференциальных уравнений

3.1. Модель оценки средней загрузки

Если $B=\infty$ (нет ограничений на пропускную способность канала – широкополосный доступ к Интернету или малое число пиров в системе), то интенсивность загрузки $D_i(t)$ можно представить в следующей форме:

$$D_i(t)=\rho\left(\sum_{j=1}^{X(t)}P_{ij}\mu+\sum_{j=1}^{Y(t)}P_{ij}\mu\right)=\rho\mu\left(\sum_{j=1}^{X(t)}P_{ij}+Y(t)\right).$$

Границы для $D_i(t)$:

$$\rho\mu Y(t)\le D_i(t)\le\rho\mu\left(Y(t)+X(t)\right).$$

Вычислим среднюю интенсивность загрузки $M(D_i(t))$ на i-м пире:

$$M(D_i(t))=\rho\mu\left(M(Y(t))+M\left(\sum_{j=1}^{X(t)}P_{ij}\right)\right).\qquad(2)$$

Границы для средней загрузки $M(D_i(t))$:

$$\rho\mu M(Y(t))\le M(D_i(t))\le\rho\mu\left(M(X(t))+M(Y(t))\right).\qquad(3)$$

Если в сети есть ограничение на максимальную пропускную способность канала (например, файл раздается большому числу пиров), т.е. $B<\infty$, то должно выполняться неравенство $D_i(t)+U_i(t)\le B$. На уровне всей сети данное ограничение имеет вид:

$$\sum_{i=1}^{N(t)}D_i(t)+\sum_{i=1}^{N(t)}U_i(t)\le\sum_{i=1}^{N(t)}B_i=BN(t).\qquad(4)$$

Запишем уравнение баланса:

$$\sum_{i=1}^{N(t)}D_i(t)=\sum_{i=1}^{N(t)}U_i(t).\qquad(5)$$

Подставляя выражение (5) в (4), получим

$$2\sum_{i=1}^{N(t)}D_i(t)\le B\,N(t).$$

Применим оператор математического ожидания к предыдущему выражению:

$$M\left(\sum_{i=1}^{N(t)}D_i(t)\right)\le\frac{B}{2}M\left(N(t)\right).\qquad(6)$$

Используя тождество Вальда ($M\left(\sum_{j=1}^{Y}X_j\right)=M(Y)M(X)$, где Y – случайная величина, а X_j – независимые, одинаково распределенные случайные величины, Y не зависит от X_j), выражение (6) можно преобразовать к следующему виду:

$$M\left(D_i(t)\right)\le\frac{B}{2}.\qquad(7)$$

Объединяя (2) и (7), можно записать общее выражение для средней интенсивности загрузки *i*-го пира:

$$M\left(D_i(t)\right)\approx\min\left\{\rho\mu\left(M(Y(t))+M\left(\sum_{j=1}^{X(t)}P_{ij}\right)\right),\frac{B}{2}\right\}.\qquad(8)$$

Объединяя выражения (3) и (8), можно записать граничные неравенства для средней загрузки:

$$\rho\mu M\left(Y(t)\right)\le M\left(D_i(t)\right)\le\min\left\{\rho\mu\left(M\left(X(t)\right)+M\left(Y(t)\right)\right),\frac{B}{2}\right\},\qquad(9)$$

которые в предположении о независимости и одинаковой распределенности интенсивностей загрузок $D_i(t)$ можно переписать в виде:

$$\rho\mu M\left(Y(t)\right)\le M\left(D(t)\right)\le\min\left\{\rho\mu\left(M\left(X(t)\right)+M\left(Y(t)\right)\right),\frac{B}{2}\right\}.\qquad(9')$$

3.2. Стохастические дифференциальные уравнения, описывающие изменения числа личеров $X(t)$ и сидеров $Y(t)$ в системе

Составим дифференциальные уравнения, описывающие изменения числа личеров $X(t)$ и сидеров $Y(t)$ в системе:

$$dX(t)=dN(t)-\frac{\sum_{i=1}^{X(t)}D_i(t)}{\sum_{i=1}^{X(t)}sc_i}dt,$$
$$dY(t)=\frac{\sum_{i=1}^{X(t)}D_i(t)}{\sum_{i=1}^{X(t)}sc_i}dt-\gamma Y(t)dt.\qquad(10)$$

Здесь: $\sum_{i=1}^{X(t)}D_i(t)$ – объем информации, переданный личерам; $\sum_{i=1}^{X(t)}sc_i$ – объем данных, содержащийся у личеров, s – размер передаваемого куска файла в битах; $dN(t)$ – интенсивность поступления новых личеров в сеть; $\frac{\sum_{i=1}^{X(t)}D_i(t)}{\sum_{i=1}^{X(t)}sc_i}dt$ – интенсивность перехода личеров в сидеры; γ – интенсивность ухода сидеров из сети.

Упростим выражение (10), предположив, что $\sum_{i=1}^{X(t)} c_i \approx M$. Тогда получим:

$$\begin{aligned} dX(t) &\approx dN(t) - \frac{\sum_{i=1}^{X(t)} D_i(t)}{sM} dt, \\ dY(t) &\approx \frac{\sum_{i=1}^{X(t)} D_i(t)}{sM} dt - \gamma Y(t) dt. \end{aligned} \qquad (10')$$

Применим оператор математического ожидания к (10'):

$$\begin{aligned} dM(X(t)) &\approx dM(N(t)) - \frac{1}{sM} M\left(\sum_{i=1}^{X(t)} D_i(t)\right) dt, \\ dM(Y(t)) &\approx \frac{1}{sM} M\left(\sum_{i=1}^{X(t)} D_i(t)\right) dt - \gamma M\big(Y(t)\big) dt. \end{aligned} \qquad (11)$$

Воспользуемся тождеством Вальда и получим систему дифференциальных выражений:

$$\begin{aligned} dM(X(t)) &\approx \lambda dt - \frac{1}{sM} M\big(D(t)\big) M\big(X(t)\big) dt, \\ dM(Y(t)) &\approx \frac{1}{sM} M\big(D(t)\big) M\big(X(t)\big) dt - \gamma M\big(Y(t)\big) dt. \end{aligned} \qquad (12)$$

Используя выражение (8) (учитываем возможные ограничения на пропускную способность соединений), получаем:

$$\begin{aligned} dM(X(t)) &\approx \lambda dt - \frac{1}{sM} \min\left\{\rho\mu\left(M(Y(t)) + M\left(\sum_{j=1}^{X(t)} P_{ij}\right)\right), \frac{B}{2}\right\} M\big(X(t)\big) dt, \\ dM(Y(t)) &\approx \frac{1}{sM} \min\left\{\rho\mu\left(M(Y(t)) + M\left(\sum_{j=1}^{X(t)} P_{ij}\right)\right), \frac{B}{2}\right\} M\big(X(t)\big) dt - \gamma M\big(Y(t)\big) dt. \end{aligned} \qquad (13)$$

С помощью выражения (9') преобразуем систему уравнений (13) в систему неравенств:

$$\begin{aligned} &\lambda dt - \frac{1}{sM} \min\left\{\rho\mu\big(M(Y(t)) + M\big(X(t)\big)\big), \frac{B}{2}\right\} M\big(X(t)\big) dt \le dM(X(t)) \le \\ &\qquad \le \lambda dt - \frac{\rho\mu}{sM} M(Y(t)) M\big(X(t)\big) dt, \\ &\frac{\rho\mu}{sM} M(Y(t)) M\big(X(t)\big) dt - \gamma M\big(Y(t)\big) dt \le dM(Y(t)) \le \\ &\qquad \le \frac{1}{sM} \min\left\{\rho\mu\big(M(Y(t)) + M\big(X(t)\big)\big), \frac{B}{2}\right\} M\big(X(t)\big) dt - \gamma M\big(Y(t)\big) dt. \end{aligned} \qquad (14)$$

3.3. Стационарный режим

Переходя к стационарному режиму, система дифференциальных неравенств (14) преобразуется в систему неравенств:

$$\begin{aligned} &\lambda - \frac{1}{sM} \min\left\{\rho\mu\big(\overline{Y} + \overline{X}\big), \frac{B}{2}\right\} \overline{X} \le 0 \le \lambda - \frac{\rho\mu}{sM} \overline{Y}\,\overline{X}, \\ &\frac{\rho\mu}{sM} \overline{Y}\,\overline{X} - \gamma \overline{Y} \le 0 \le \frac{1}{sM} \min\left\{\rho\mu\big(\overline{Y} + \overline{X}\big), \frac{B}{2}\right\} \overline{X} - \gamma \overline{Y}. \end{aligned} \qquad (15)$$

Здесь $\overline{Y} = M\big(Y(t)\big)$, а $\overline{X} = M\big(X(t)\big)$.

Введя обозначения $\alpha = \frac{\rho\mu}{sM}$ и $\beta = \frac{B}{2\mu\rho}$, систему (15) можно переписать в виде:

$$\begin{aligned} &\lambda - \alpha \min\left\{\overline{Y} + \overline{X}, \beta\right\} \overline{X} \le 0 \le \lambda - \alpha \overline{Y}\,\overline{X}, \\ &\alpha \overline{Y}\,\overline{X} - \gamma \overline{Y} \le 0 \le \alpha \min\left\{\overline{Y} + \overline{X}, \beta\right\} \overline{X} - \gamma \overline{Y}. \end{aligned} \qquad (15')$$

Решим систему неравенств (15'). Для этого рассмотрим два случая: 1) $\overline{Y} + \overline{X} < \beta$, т.е. нет ограничений на пропускную способность соединения или мало пиров в сети и 2) $\beta < \overline{Y} + \overline{X}$ (есть ограничения на пропускную способность соединений или много пиров в сети).

3.4. Случай отсутствия ограничений на пропускную способность соединения.

Рассмотрим случай, когда $\overline{Y} + \overline{X} < \beta$, т.е. нет ограничений на пропускную способность соединения или мало пиров в сети.

Система неравенств (15') преобразуется в следующую систему:

$$\begin{cases} \lambda - \alpha\big(\overline{Y} + \overline{X}\big)\overline{X} \le 0 \le \lambda - \alpha \overline{Y}\,\overline{X}, \\ \alpha \overline{Y}\,\overline{X} - \gamma \overline{Y} \le 0 \le \alpha\big(\overline{Y} + \overline{X}\big)\overline{X} - \gamma \overline{Y}, \end{cases}$$

которая, в свою очередь, записывается как две системы:

$$\begin{cases} \lambda - \alpha\big(\overline{Y} + \overline{X}\big)\overline{X} \le 0, \\ \alpha \overline{Y}\,\overline{X} - \gamma \overline{Y} \le 0, \end{cases} \text{ и } \begin{cases} \lambda - \alpha \overline{Y}\,\overline{X} \ge 0, \\ \alpha\big(\overline{Y} + \overline{X}\big)\overline{X} - \gamma \overline{Y} \ge 0. \end{cases}$$

Решая эти системы, получим выражения для нахождения границ средних значений числа личеров и сидеров соответственно:

$$\begin{cases} \overline{X} \le \dfrac{\gamma}{\alpha}, \\ \overline{Y} \ge \dfrac{\lambda}{\gamma}, \end{cases} \text{ и } \begin{cases} \overline{Y} \le \dfrac{\lambda}{\alpha \overline{X}}, \\ \alpha^2 \overline{X}^3 + \lambda\alpha \overline{X} - \gamma\lambda \ge 0. \end{cases}$$

3.5. Случай наличия ограничения на пропускную способность соединений

Пусть теперь $\overline{Y} + \overline{X} > \beta$, т.е. существуют ограничения на пропускную способность соединений или много пиров в сети.

Система неравенств (15') преобразуется в следующую систему:

$$\begin{cases} \lambda - \alpha\beta \overline{X} \le 0 \le \lambda - \alpha \overline{Y}\,\overline{X}, \\ \alpha \overline{Y}\,\overline{X} - \gamma \overline{Y} \le 0 \le \beta \overline{X} - \gamma \overline{Y}, \end{cases}$$

которая снова представляется в виде двух систем:

$$\begin{cases} \lambda - \alpha\beta \overline{X} \le 0, \\ \alpha \overline{Y}\,\overline{X} - \gamma \overline{Y} \le 0, \end{cases} \text{ и } \begin{cases} 0 \le \lambda - \alpha \overline{Y}\,\overline{X}, \\ 0 \le \beta \overline{X} - \gamma \overline{Y}. \end{cases}$$

Решая системы, получим границы для средних значений числа личеров и сидеров соответственно:

$$\begin{cases} \max\left\{\dfrac{\lambda}{\alpha\beta}, \sqrt{\dfrac{\lambda\gamma}{\alpha\beta}}\right\} \le \overline{X} \le \dfrac{\gamma}{\alpha}, \\ \overline{Y} \le \sqrt{\dfrac{\lambda\beta}{\alpha\gamma}}. \end{cases}$$

Используя теорему Литтла, можно получить среднее время загрузки файла $\overline{T}_d$ как $\overline{T}_d = \frac{\overline{X}}{\lambda}$, а точнее говоря, границы изменения среднего времени загрузки файла личером.

4. Заключение

Данная работа является лишь началом исследования авторов по данной тематике, поэтому в некоторых разделах сформулированы лишь начальные результаты.

С помощью аппарата стохастических дифференциальных уравнений построена математическая модель распространения данных по протоколу типа BitTorrent, описывающая процесс распространения на большое число компьютеров (рабочих станций пользователей) дисплейного класса заранее подготовленного файла образа операционной системы и программного обеспечения.

Хотя полученная модель не ограничена по количеству хостов, она вполне адекватно может описывать задачу с ограниченным количеством узлов в предположении, что стационарный режим достигается прежде, чем будут исчерпаны личеры. Это предположение может работать, например, в случае, если наличествующее число хостов выбирает всю пропускную способность сети.

Получены аналитические выражения для границ среднего числа личеров, сидеров, а также выражения средней пропускной способности системы и времени загрузки данных.

Литература

1. *А. Адаму, Ю.В. Гайдамака*. Анализ вероятности непрерывного воспроизведения видеопотока в P 2 P -сети // «Вестник РУДН. Серия «Математика. Информатика. Физика».» – М.: Изд-во РУДН. – 2011. – №4. - С. 38-46.
2. *Aminu Adamu, Yuliya Gaidamaka, and Andrey Samuylov* Discrete Markov Chain Model for Analyzing Probability Measures of P2P Streaming Network // Lecture Notes in Computer Science. – Germany, Heidelberg: Springer. – 2011. – Vol. 6869. Pp. 428-439.
3. *B. Fan, D.-M. Chiu, and J. Lui,* "Stochastic differential equation approach to model bittorrent-like file sharing systems," in Proc. of 14th IEEE International Workshop on Quality of Service, 2006
4. *L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang*, "A performance study of bittorrent-like peer-to-peer systems," in IEEE Journal on Selected Areas in Communications, vol. 25, no. 1, January 2007
5. *Y. Tian, D. Wu, and K. Ng,* "Modeling, analysis and improvement for bittorrent-like file sharing networks," in Proc. of IEEE INFOCOM, 2006.
6. *Wei-Cherng Liao, Fragkiskos Papadopoulos, Konstantinos Psounis* Performance Analysis of BitTorrent-like Systems with Heterogeneous Users http://www-scf.usc.edu/~weicherl/paper/bttech_final.pdf.
7. *B. Fan, D.-M. Chiu, and J. C. S. Lui*, "Performance modeling of bittorrent-like file sharing systems," Dept. of Computer Sci. and Eng. , the Chinese University of Hongkong, Tech. Rep., Nov 2005.
8. *D. Qiu and R. Srikant*, "Modeling and performance analysis of bittorrent-like peer-to-peer networks," in *Proc. ACM SIGCOMM*, 2004.
9. *X. Yang and G. de Veciana*, "Service capacity of peer to peer networks," in *Proceedings of IEEE INFOCOM*, 2004.
10. *R. Brockett*, "Stochastic control," Harvard University, lecture notes.

The stochastic model of Bittorent protocol

Kulyabov D.S., *PhD, associate professor, telecommunication systems department, PFUR, dskulyabov@gmail.com*
Korolkova A.V., *PhD, associate professor, telecommunication systems department, PFUR, akorolkova@sci.pfu.edu.ru*
Zaryadov I.S., *PhD, assistant professor, probability theory and mathematical statistics department, PFUR, izaryadov@gmail.com*

Abstract

Paradigms of centralized and peer-to-peer communications historically follow each other with the development of the computer equipment and telecommunications networks. The peer-to-peer methods of data transmission that enable the movement of large amounts of data without expanding the capacity of existing transmission lines are popular now. The BitTorrent protocol is one of the most common protocols of peer-to-peer networks. The main task of the protocol is to distribute the file to the greatest number of peers without reducing the file accessibility. In this paper the mathematical model of data transmission by BitTorrent-like protocol based on stochastic differential equations is presented. The model describes the distribution of operation system and software image file among a high number of computers (workstations) in a display class. The proposed model allows analyze such main protocol characteristics as the mean value of leechers and seeders in the system, average throughput of the system and data loading time. The last characteristic is very important for the analysis of one of the approaches of solving one of the main tasks of system administration - the automation of the process of operation system and software installation on high number workplaces in the limited time - that is the advantage of the BitTorrent protocol over other possible solutions of the problem.

Keywords: *BitTorrent protocol, peer-to-peer network, mathematical model, system of stochastic differential equations, means data download time.*