



УДК 004.94

Вьет Нгуен Дык, А. С. Тамазян
Санкт-Петербургский государственный электротехнический
университет "ЛЭТИ" им. В. И. Ульянова (Ленина)

Модель сетевого трафика на основе суперпозиции однородных потоков пользовательских запросов

Рассмотрено выделение пользовательских сессий из сетевого трафика. Выполнен статистический анализ межсессионных интервалов в канале инфокоммуникационной сети. Предложено описание распределений межсессионных интервалов q -экспоненциальным распределением. Предложена модель агрегированного трафика инфокоммуникационной сети на основе суперпозиции потоков данных коротких по длительности фрагментов трафика и доказана ее применимость на примере эмпирических данных.

Инфокоммуникационная сеть, сетевой канал, сетевой трафик, моделирование, суперпозиция

Понимание принципов, лежащих в основе сложной динамики пользовательской активности в инфокоммуникационных сетях, является одной из наиболее актуальных задач при их разработке и эксплуатации. Ранее было показано [1], что агрегированный сетевой трафик обладает сильной нелинейностью, в связи с чем традиционная пуассоновская модель, основанная на случайной динамике обращения к информации, в данном случае неприменима.

В настоящей статье проанализированы данные, полученные из двух источников. Первый источник – записи трафика канала опорной сети WIDE, соединяющей большое число университетов и научно-исследовательских центров Японии с вышестоящим интернет-провайдером (sample-point-F), собранные в рамках WIDE Project, свободно распространяемые в сети для научно-исследовательских целей [2]. Для анализа были выбраны данные за 9 полных дней, а именно записи трафика с 18.03.2008 по 20.03.2008, с 30.09.2009 по 04.02.2009 и с 13.04.2010 по 14.04.2010 гг. Второй набор данных был собран с канала, соединяющего локальную сеть общежития Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" с интернет-провайдером. Как и в первом случае, было отобрано 9 записей трафика, а именно данные с 17.03.2015 по 18.03.2015, с 16.04.2015 по 19.04.2015 и с 26.04.2016 по 28.04.2016 гг.

Исходные данные трафика были обработаны программой "tshark" для получения списка TCP-пакетов, переданных через канал, включая время прибытия пакета, IP-адрес пользователя–получателя и размер пакета. Затем они были отсортированы по IP-адресам получателей, что позволило выделить пакеты, связанные с активностью каждого пользователя. В каждой такой группе пакеты сгруппировывались с межпакетным интервалом менее 1 с, что ниже типичного времени реакции конечного пользователя, в единую сессию. Межпакетный интервал, превышающий 1 с, считался разрывом между двумя отдельными сессиями.

Каждая сессия описывалась временем начала, числом пакетов в ней и объемом переданной информации, равным сумме объемов всех пакетов, переданных за время этой сессии. Таким образом, данные трафика были переведены с пакетного на сессионный уровень. Данная процедура применяется при анализе активности пользователей в сети для исключения влияния технических факторов организации сети [3]. На сессионном уровне также возможно сравнение статистических характеристик трафика инфокоммуникационных сетей различного масштаба.

Для большинства каналов инфокоммуникационных сетей в динамике их трафика ярко выражен дневной тренд, вызванный повышенной активностью во время рабочих и вечерних часов и пониженной активностью в ночные часы. Один

из распространенных методов работы с дневными трендами состоит в том, чтобы описать их отдельно от флуктуаций трафика детерминированной моделью. В агрегированном трафике форма тренда может быть оценена усреднением по кратковременным фрагментам, взятым в одно и то же время за различные дни [4]. Другой метод – анализ статистических свойств трафика для каждого кратковременного фрагмента отдельно. В этом случае длительность фрагмента трафика должна быть выбрана таким образом, чтобы в большинстве случаев его динамика в одном фрагменте значительно не изменялась. Тогда в первом приближении можно пренебречь флуктуациями интенсивности поступления пользовательских запросов в одном фрагменте и охарактеризовать каждый фрагмент средней интенсивностью поступления пользовательских запросов. Так как сессионный интервал принят превышающим 1 с, фрагменты также не должны быть слишком короткими, чтобы в них всегда было достаточное количество пользовательских сессий для надежного статистического анализа. Наконец, за время суточного цикла нужно иметь количество фрагментов, достаточное для получения хорошей статистики в течение дня, чтобы можно было наблюдать возможные различия в данных трафика за разные дни. Исходя из изложенного, а также

из эмпирических наблюдений, длительность фрагмента трафика принималась $T = 15$ мин.

На рис. 1 маркерами показаны плотности вероятности нормированных межсессионных интервалов $P(\tau/\bar{\tau}) = \bar{\tau}P(\tau)$, полученные для 15-минутных фрагментов в шести репрезентативных наборах данных локальной сети общежития (а) и канала опорной сети WIDE (б).

Из рис. 1 следует, что большинство плотностей могут быть описаны убывающей экспоненциальной функцией $P(\tau/\bar{\tau}) = \exp(-\tau/\bar{\tau})$ (серые линии). В этом случае динамика агрегированного трафика в каждом фрагменте характеризуется единственным параметром – интенсивностью поступления запросов $\beta = 1/\bar{\tau}$. На рис. 2 представлены плотности интенсивности поступления запросов $P(\beta/\bar{\beta}) = \bar{\beta}P(\beta)$ для разных суток наблюдения трафика в локальной сети общежития (а) и канала опорной сети WIDE (б). Плотность вероятности $P(\beta/\bar{\beta})$ для каждой рассматриваемой сети имеет характерную форму, мало изменяющуюся от выборки к выборке. Наблюдаемые распределения асимптотически убывают по экспоненте квадрата аргумента, поэтому их затруднительно описать широко распространенными законами распределений. Однако подобное асимптотиче-

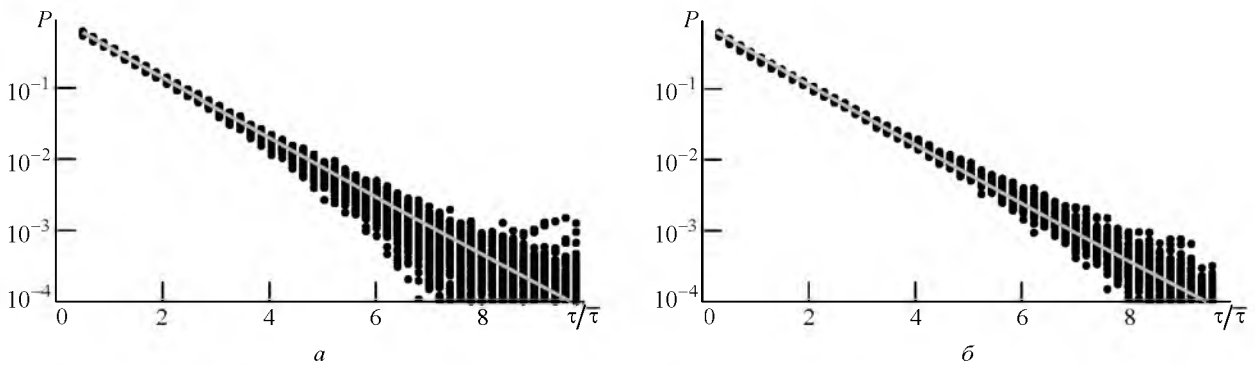


Рис. 1

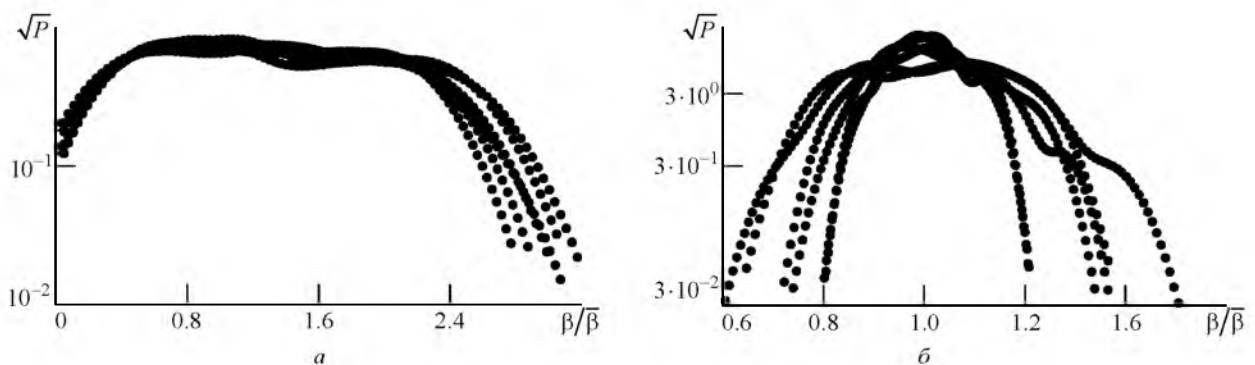


Рис. 2

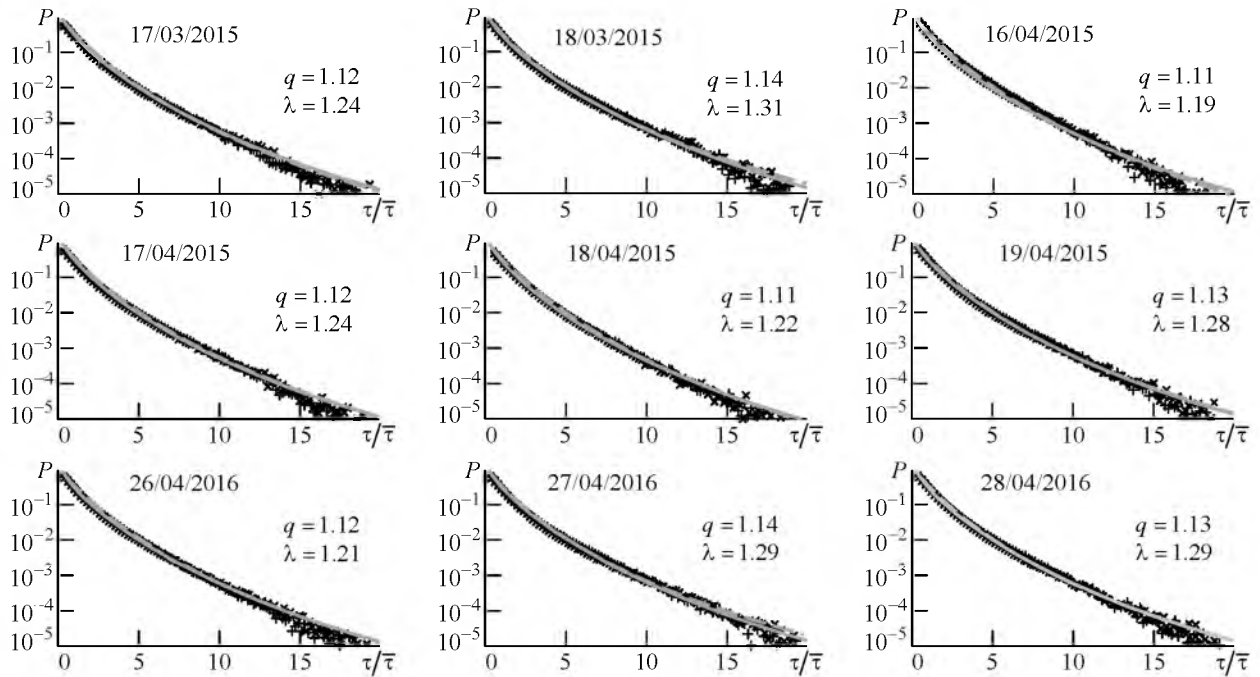


Рис. 3

ское поведение наблюдается для нормального или сходящихся к нему распределений при большом количестве степеней свободы k : гамма, биномиального, χ -квадрат и подобных. Так как рассматриваемые плотности нормированы, возможно рассчитать коэффициент формы (коэффициент вариации) $\rho = \sigma(\beta)/\bar{\beta}$. Для данных, рассмотренных в настоящей статье, получены значения $\rho = 0.35$ для каналов локальной сети общежития и $\rho = 0.1$ для опорной сети WIDE.

В дальнейшем для упрощения сравнения сетей с различной интенсивностью пользовательских запросов примем упрощенные выражения нормированных распределений $P(\beta) = \bar{\beta}P(\beta/\bar{\beta})$ и $P(\tau) = \bar{\tau}P(\tau/\bar{\tau})$.

Распределение межсессионных интервалов для агрегированного трафика может быть представлено как суперпозиция распределений множества фрагментов, описываемых плотностью вероятности экспоненциального распределения $P(\tau|\beta)$ с собственными интенсивностями β для каждого фрагмента. Так как β – случайная переменная, общее распределение межсессионных интервалов для агрегированного трафика выражается формулой полной вероятности

$$P(\tau) = \int_0^{\infty} P(\beta)P(\tau|\beta) d\beta = \int_0^{\infty} P(\beta)\beta \exp(-\beta\tau) d\beta. \quad (1)$$

В этом случае $P(\tau)$ имеет вид q -экспоненциального распределения [5]

$$P(\tau) = [1 + \lambda(q-1)\tau]^{-1/(1-q)},$$

где $q = 1 + 2/(k+2)$, а интенсивность β распределена по закону χ -квадрат с числом степеней свободы k [6]. При больших значениях k плотность вероятности этого распределения убывает по квадрату экспоненциальной функции, что наблюдается и для эмпирических данных. Коэффициент вариации составляет $\rho = \sqrt{2k}/k$. Тогда для каналов локальной сети общежития имеем $k \approx 16$ и $q \approx 1.11$, для опорной сети WIDE $k \approx 200$, $q \approx 1.01$. Отметим, что q -экспоненциальное распределение наблюдается и в межпакетных интервалах в опорной сети WIDE [7].

Для проверки сформулированных предположений проанализированы оценки плотности вероятности эмпирических нормированных межсессионных интервалов $P(\tau/\bar{\tau})$ агрегированного трафика за сутки. Полученные зависимости для каналов локальной сети общежития представлены на рис. 3, для опорной сети WIDE – на рис. 4. Здесь маркерами вида "x" показаны эмпирические значения межсессионных интервалов, а маркерами вида "+" – их значения, предсказанные моделью. Из зависимостей на рис. 3 и 4 следует, что q -экспоненциальное распределение (линии) с высокой степенью точности описывает межсессионные интервалы со значениями q , близкими к значениям,

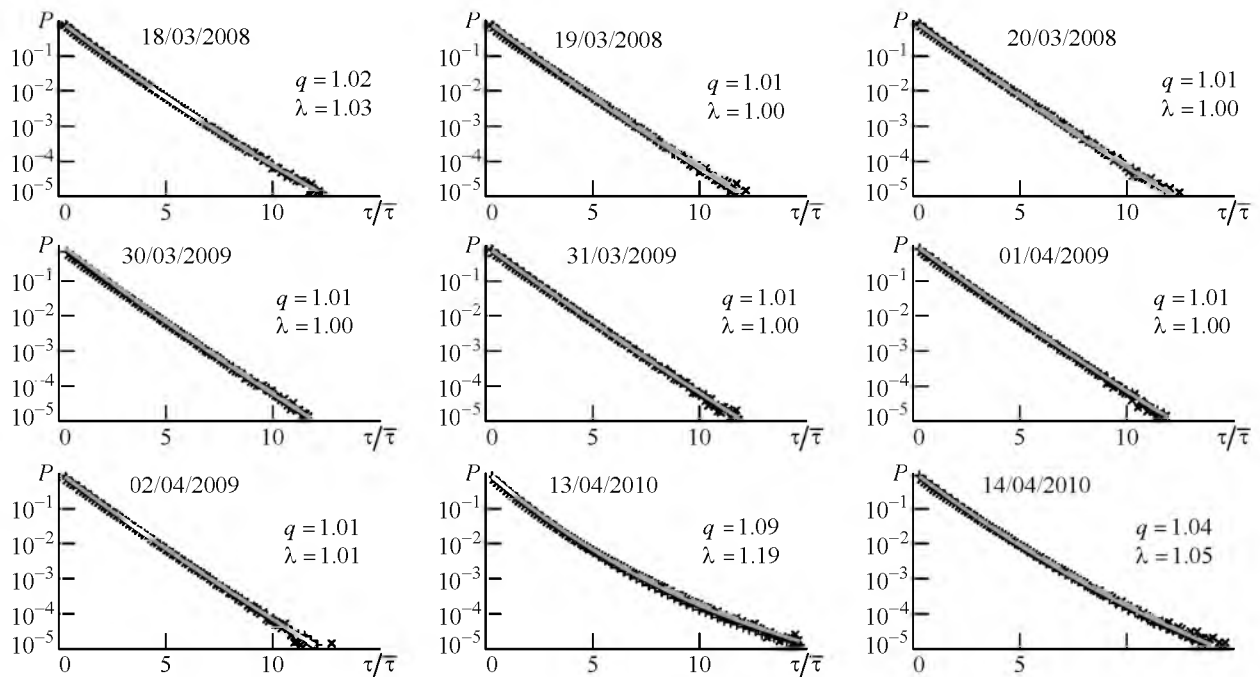


Рис. 4

предсказанной моделью на основе суперпозиции одиночных потоков пользовательских запросов.

Отклонения модельной оценки q от полученной на основании эмпирических данных могут быть объяснены отклонениями распределений эмпирических данных от нормального закона, использованного при аппроксимации, носящих случайный характер (см. рис. 2). При подстановке эмпирических распределений ежедневного трафика $P(\beta)$ в (1) и интегрировании получается достаточно точная аппроксимация $P(\tau/\bar{\tau})$ для всех дней (линии) [8].

Моделирование потоков запросов суперпозицией пуассоновских потоков с различными локальными интенсивностями на 15-минутных временных интервалах приводит к аналогичному q -экспоненциальному виду плотности распределения вероятностей $P(\tau/\bar{\tau})$, получаемой при численном интегрировании согласно (1).

Авторы выражают благодарность О. Ю. Худякову за помощь в сборе данных трафика, использованных при написании статьи.

СПИСОК ЛИТЕРАТУРЫ

1. On The Self-Similar Nature of Ethernet Traffic / W. E. Leland, W. Willinger, M. S. Taqqu, D. V. Wilson // IEEE/ACM Trans. on Netw. 1994. Vol. 2, no. 1. P. 1–15.
2. URL: <http://mawi.wide.ad.jp/mawi> (дата обращения: 25.01.2017).
3. A Compound Model for TCP Connection Arrivals for LAN and WAN Applications / C. Nuzman, I. Saniee, W. Sweldens, A. Weiss // Computer Networks. 2002. Vol. 40, № 3. P. 319–337.
4. A Novel Network Traffic Anomaly Detection Model Based on Superstatistics Theory / Y. Yang, H. Hu, W. Xiong, F. Ding // J. Networks. 2011. Vol. 6, № 2. P. 311–318.
5. Beck C., Cohen E. D. G., Superstatistics // Phys. A. 2003. Vol. 322. P. 267–275.
6. Briggs K., Beck C. Modelling Train Delays with q -exponential Functions // Phys. A. 2007. Vol. 378, iss. 2. P. 498–504.
7. Tamazian A., Markelov O. A., Bogachev M. I. Traffic Analysis on the WIDE Backbone Link: from Transport Level to End User Activity // IEEE NW Russia Young Res. in Electric and Electronic Eng. Conf., Saint Petersburg, 2–3 Febr. 2016. P. 356–359. URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7448192> (дата обращения: 26.01.2017).
8. Universal Model for Collective Access Patterns in the Internet Traffic Dynamics: A Superstatistical Approach / A. Tamazian, V. D. Nguyen, O. A. Markelov, M. I. Bogachev // Euro Phys. Lett. 2016. Vol. 115, № 1. P. 10008.

Статья поступила в редакцию 14 ноября 2016 г.

Для цитирования: Вьет Нгуен Дык, Тамазян А. С. Модель сетевого трафика на основе суперпозиции однородных потоков пользовательских запросов // Изв. вузов России. Радиоэлектроника. 2017. № 1. С. 40–44.

Вьет Нгуен Дык – аспирант по кафедре радиотехнических систем Санкт-Петербургского государственного электротехнического университета "ЛЭТИ". Окончил Технический университет г. Ханой (2010) по специальности "Радиоэлектронные и телекоммуникационные системы". Автор четырех научных публикаций. Сфера научных интересов – телекоммуникационные и инфокоммуникационные системы; математическое моделирование; системы массового обслуживания.

E-mail: ndvietleti@gmail.com

Тамазян Араик Симакович – магистр техники и технологии по направлению "Телекоммуникации" (2014), инженер II категории кафедры радиотехнических систем Санкт-Петербургского государственного электротехнического университета "ЛЭТИ". Автор 12 научных публикаций. Сфера научных интересов – статистический анализ и моделирование инфокоммуникационных систем и систем массового обслуживания.

E-mail: araik.tamazian@outlook.com

Viet Nguyen Duc, A. S. Tamazian
Saint Petersburg Electrotechnical University "LETI"

Network Traffic Model Based on Superposition of Single User Requests Flows

Abstract. User session extraction from network traffic data is considered. Statistical analysis of inter-session times of infocommunication network link was performed. Q-exponential distribution was proposed to describe the inter-session times distribution. Aggregated traffic model of infocommunication network link based on superposition of short-term traffic fragments data was suggested. The aggregated traffic model was verified by means of empirical traffic data.

Key words: Infocommunication Network, Network Link, Network Traffic, Modeling, Superposition

REFERENCES

1. Leland W. E., Willinger W., Taqqu M. S., Wilson D. V. On The Self-Similar Nature of Ethernet Traffic. IEEE, ACM Trans. on Netw., 1994, vol. 2, no. 1, pp. 1–15.
2. Available at: <http://mawi.wide.ad.jp/mawi> (accessed 25 January 2017).
3. Nuzman C., Saniee I., Sweldens W., Weiss A. A. Compound Model for TCP Connection Arrivals for LAN and WAN Applications. Computer Networks, 2002, vol. 40, no. 3, pp. 319–337.
4. Yang Y., Hu H., Xiong W., Ding F. A Novel Network Traffic Anomaly Detection Model Based on Superstatistics Theory. J. Networks, 2011, vol. 6, no. 2, pp. 311–318.
5. Beck C., Cohen E. D. G., Superstatistics. Physica A. 2003, vol. 322, pp. 267–275.
6. Briggs K., Beck C., Modelling Train Delays with q-exponential Functions. Physica A. 2007, vol. 378, no. 2, pp. 498–504.
7. Tamazian A., Markelov O. A., Bogachev M. I. Traffic Analysis on the WIDE Backbone Link: from Transport Level to End User Activity. IEEE NW Russia Young Res. in Electric and Electronic Eng. Conf., Saint Petersburg, 2–3 Febr. 2016, pp. 356–359. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7448192> (accessed 26 January 2017).
8. Tamazian A., Nguyen V. D., Markelov O. A., Bogachev M. I. Universal Model for Collective Access Patterns in the Internet Traffic Dynamics: A Superstatistical Approach. EPL, 2016, vol. 115, no. 1, p. 10008.

Received November, 4, 2016

For citation: Duc Viet Nguyen, Tamazian A. S. Network Traffic Model Based on Superposition of Single User Requests Flows. *Izvestiya Vysshikh Uchebnykh Zavedenii Rossii. Radioelektronika* [Journal of the Russian Universities. Radioelectronics]. 2017, no. 1, pp. 40–44. (In Russian).

Viet Nguyen Duc – Dipl.-engineer on radio electronic and telecommunication systems (2010, Hanoi University of Science and Technology), post graduate student of the Department of Radio Equipment Systems of Saint Petersburg Electrotechnical University "LETI". The author of four scientific publications. Area of expertise: telecommunication and infocommunication systems; mathematical modelling; queueing systems.

E-mail: ndvietleti@gmail.com

Araik S. Tamazian – Master's Degree in telecommunications (2014), 2nd category engineer of the Department of Radio Equipment Systems of Saint Petersburg Electrotechnical University "LETI". The author of 12 scientific publications. Area of expertise: statistical analysis and modelling of the infocommunication and queueing systems.

E-mail: araik.tamazian@outlook.com