

МОДЕЛЬ УПРАВЛЕНИЯ ХРАНЕНИЕМ ТРАФИКА ДАННЫХ

Е. Д. ПОЙМАНОВА, Т. М. ТАТАРНИКОВА, Е. В. КРАЕВА

*Российский государственный гидрометеорологический университет,
192007, Санкт-Петербург, Россия
E-mail: e.d.poymanova@gmail.com*

Представлено комплексное решение для управления записью и хранением трафика в системах хранения данных. Предложена структура программно-аппаратного комплекса управления физическим хранилищем данных для систем, которые могут использоваться владельцами технологических сетей связи для хранения трафика. Рассмотрены механизмы управления, такие как распределение данных по различным носителям с использованием нейронных сетей Кохонена и прогнозирование увеличения емкости хранилища с помощью статистической модели и методов машинного обучения.

Ключевые слова: трафик, система хранения данных, распределение данных, физическое хранилище данных, машинное обучение, нейронная сеть, прогнозирование

Введение. Согласно требованиям современного законодательства в сфере безопасности граждан перед различными организациями и, в частности, владельцами технологических сетей связи стоят серьезные задачи в области хранения данных.

Антитеррористические поправки, принятые в 2016 г., обязывают операторов связи хранить метаданные трафика в течение трех лет, а сам трафик — в течение шести месяцев. Кроме того, в 2020 г. Министерством цифрового развития, связи и массовых коммуникаций РФ был внесен законопроект, согласно которому необходимо хранить трафик в течение трех лет [1].

Также существует законодательная норма, обязывающая ежегодно увеличивать объем хранилищ трафика на 15 %. За 2016—2019 гг. объем интернет-трафика составил от 32470,78 до 61 226,22 Пбайт (рис. 1) [2], т.е. всего за четыре года объем трафика вырос почти вдвое. Следовательно, норма в 15 % по ежегодному увеличению емкости хранилищ является недостаточной, при этом очевидно, что ее реализация требует значительных финансовых затрат. Как пример, в таблице [3] приведены затраты операторов связи РФ на закупку оборудования для систем хранения данных (СХД). С другой стороны, современные информационные технологии позволяют управлять ресурсами СХД, эффективно их использовать и, как следствие, избегать излишних затрат.

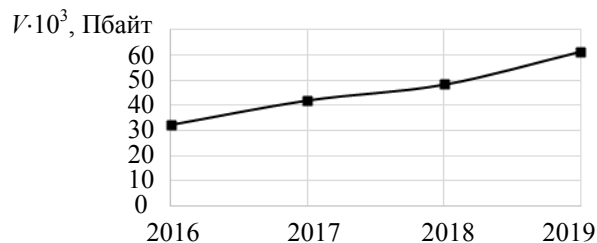


Рис. 1

Год, месяц	Оператор связи	Стоимость, руб.
2018, май	Мегафон	923,5 млн
2019, март	МТС	14 млрд
2020, октябрь	Ростелеком	7,8 млрд
Запланировано	Tele2	4,3 млрд

Так, система хранения данных может управлять записью входящего потока данных и, во-первых, распределять его по разным типам носителей, а во-вторых, отслеживать состояние хранилища и составлять прогноз своевременного наращивания его емкости [4, 5].

Управление физическим хранилищем данных при записи и хранении трафика. Авторами было проведено исследование, в котором СХД рассматривается как система управления хранением, выполняющая следующие функции:

- распределение файлов данных по различным типам носителей в зависимости от размера файла и времени хранения;
- мониторинг состояния хранилища на основе срезов состояния;
- прогноз наращивания емкости хранилища [6, 7].

Схема системы управления хранением данных представлена на рис. 2.

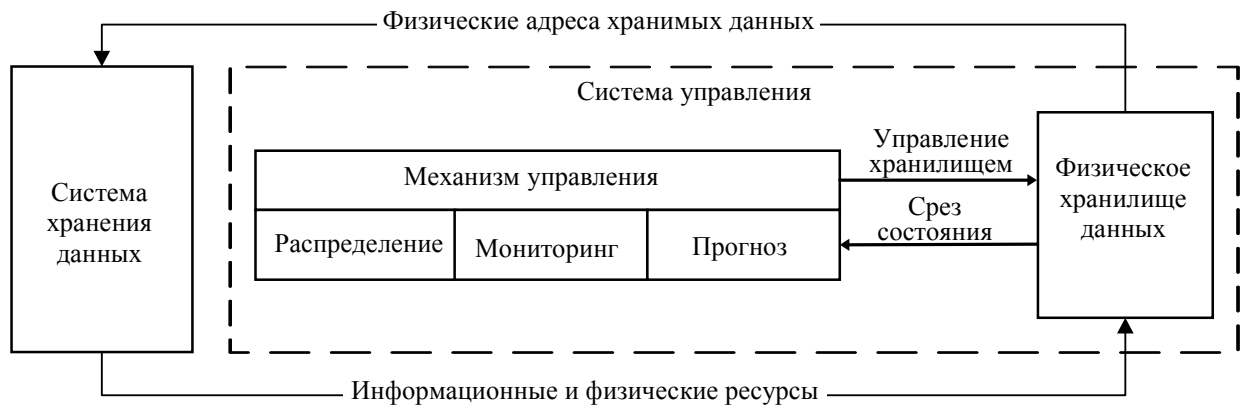


Рис. 2

Очевидно, что для реализации такой системы необходим программно-аппаратный комплекс, выполняющий указанные выше функции. В состав комплекса, структура которого представлена на рис. 3, предлагается включить программируемый логический контроллер, осуществляющий распределение файлов по носителям, и программное обеспечение, осуществляющее мониторинг состояния физического хранилища и построение прогноза наращивания его объема.

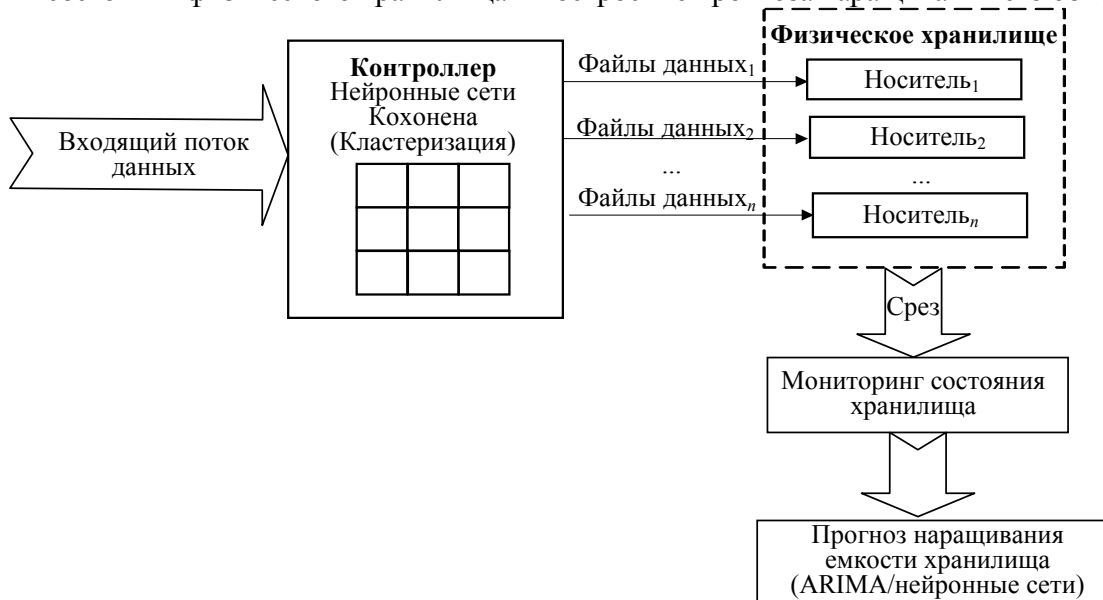


Рис. 3

Предполагается, что на контроллер поступает входящий поток данных (например, интернет-трафик). Контроллер осуществляет кластеризацию входящего трафика с помощью нейронных сетей Кохонена и в соответствии с полученной топологической картой распределяет файлы данных по носителям в физическом хранилище.

Физическое хранилище может быть организовано в соответствии с характеристиками записываемой информации. В работе [8] было рассмотрено матричное хранилище размерностью 3×3 в предположении, что распределение файлов осуществляется сначала по одному из уровней хранения в зависимости от времени хранения, а затем по носителям внутри уровня в зависимости от размера файла.

Это решение легко адаптируется к потребностям владельцев технологических сетей связи [9]. Поскольку время хранения файлов данных, а также метаданных ограничено тремя годами (как указано выше), файлы данных можно распределять по различным типам носителей в зависимости от типа данных (текст, звук, видео) и от размера.

Структура физического хранилища данных (рис. 4) определяется администратором системы хранения и может содержать, например, RAID-массивы для текстовых файлов, стримеры для аудио- и видеофайлов. Кроме того, тома внутри RAID-массива могут иметь разные файловые системы с различными размерами логического блока данных, что позволит избежать „недозаполненности“ файлов при записи и, следовательно, сэкономить дисковое пространство на RAID-массивах. Сравнение файлов, содержащих одинаковое количество данных, но имеющих логические блоки разного размера, отражено на рис. 5.

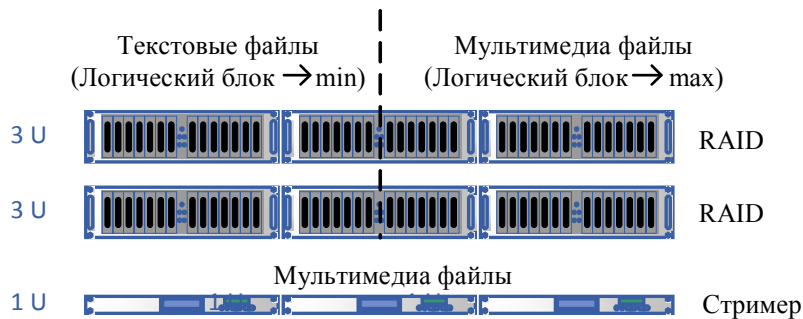


Рис. 4

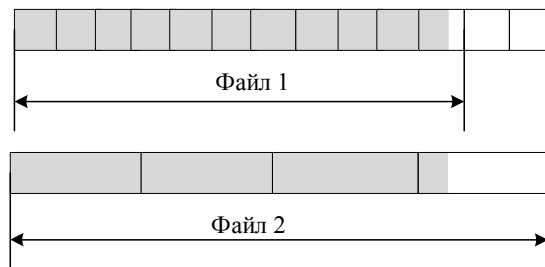


Рис. 5

Мониторинг емкости хранилища можно осуществлять на основе данных о его состоянии в определенный момент времени (срезы состояния), поступающих с физического хранилища данных. Эти срезы состояния должны показывать заполненность каждого носителя (тома носителя) данных в физическом хранилище. Используя данные мониторинга, можно строить прогноз наращивания емкости физического хранилища. При этом рассматривается каждый уровень хранения, содержащий определенный тип носителей. Прогноз наращивания емкости основан на модели, представленной на рис. 6 [10], где V_{cur} — текущий объем,

$$V_{\text{lim}} = \int_1^{t_{\text{lim}}} f(t) dt = T \sum_1^{t_{\text{lim}}} f(t); \quad V_{\text{max}} = \int_1^{t_{\text{max}}} f(t) dt = T \sum_1^{t_{\text{max}}} f(t),$$

где t_{lim} и t_{max} — время достижения предельной и максимальной емкости носителя соответственно; $f(t)$ — функция входящего потока данных; T — шаг, равный единице выбранной шкалы времени.

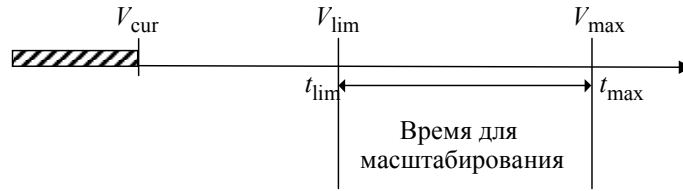


Рис. 6

Задача прогнозирования сводится к определению времени достижения предельной и максимальной емкости каждого носителя [7].

Для решения поставленной задачи необходимо спрогнозировать количество входящего в СХД трафика.

Прогноз может быть осуществлен различными методами, при этом следует учитывать особенности потока данных, поступающего на запись. Вследствие неравномерной активности пользователей (связанной с выходными и рабочими днями, периодами отпусков и т.д.) входящий поток данных (LTE) является неоднородным и имеет сезонную структуру. В качестве примера на рис. 7 приведен график входящего потока LTE-трафика компании MTS.

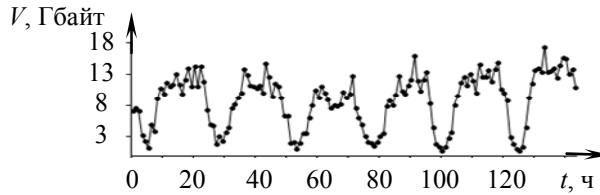


Рис. 7

В работе [11] был проведен сравнительный анализ различных методов прогнозирования — статистического прогнозирования с использованием модели авторегрессии и интегрированного скользящего среднего (ARIMA) и методов машинного обучения. Результаты показали, что для краткосрочных прогнозов наиболее подходящей является модель ARIMA (рис. 8), а для среднесрочных прогнозов — методы машинного обучения (рис. 9: а — дерево решений, б — случайный лес)*; на всех графиках кривая 1 — исходный трафик, 2 — прогноз.

Для автоматизации процесса прогнозирования наращивания емкости было разработано приложение, позволяющее строить прогноз для каждой ячейки матрицы хранения, т.е. для каждого носителя или тома носителя физического хранилища данных [12].

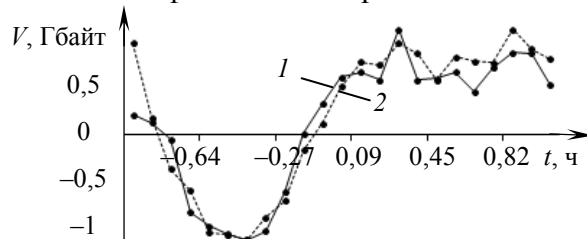


Рис. 8

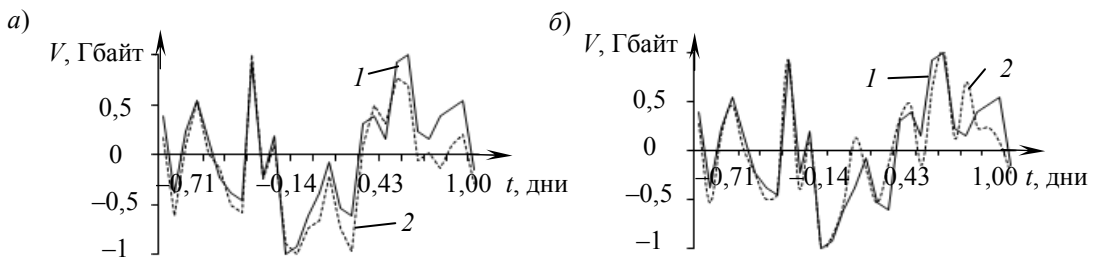


Рис. 9

* На рис. 8, 9 по осям абсцисс приведены нормализованные данные.

Таким образом, для дальнейшей реализации программно-аппаратного комплекса необходимо разработать программируемый логический контроллер, распределяющий файлы внутри хранилища физических данных на основе разработанных моделей.

Программируемые логические контроллеры нашли широкое применение в системах автоматического управления. Производительность современных контроллеров позволяет применять в них наиболее эффективные алгоритмы управления, такие, например, как нейронные сети.

Заключение. Нормы современного законодательства обязывают владельцев технологических сетей связи хранить большой объем данных, используя собственные СХД. Это приводит к значительным экономическим затратам, которые в результате ложатся на конечного потребителя телекоммуникационных услуг.

В то же время современные технологии позволяют создавать системы управления физическим хранилищем данных, которые могут эффективно расходовать физические ресурсы СХД. Существующие технологии виртуализации позволяют создавать структуры, содержащие различные типы носителей данных, и распределять по ним файлы сохраненного трафика в зависимости от заранее определенных характеристик файлов.

Поскольку существует необходимость в регулярном масштабировании хранилища данных, необходимо отслеживать его состояние и масштабировать только те носители, предельные показатели емкости которых стремятся к максимальным значениям. Прогнозирование увеличения емкости позволяет производить масштабирование своевременно.

Таким образом, разделение общего входящего потока данных по носителям, прогнозирование потребления емкости и мониторинг состояния физического хранилища данных позволяют владельцам технологических сетей связи рационально использовать физические ресурсы СХД и избегать неоправданных расходов при увеличении хранилища.

СПИСОК ЛИТЕРАТУРЫ

1. <https://tass.ru/ekonomika/9574021>
2. <https://digital.gov.ru/ru/pages/statistika-otrasli/>
3. <https://www.kommersant.ru/doc/4522028>
4. *Farley M.* Building Storage Networks. McGraw-Hall: Osborne, 2001.
5. Information Storage and Management. John Wiley & Sons Inc., 2016.
6. *Tatarnikova T. M., Poymanova E. D.* Algorithms for placing files in tiered storage using Kohonen map // CEUR Workshop Proc. ISTMC'2019 Selected Papers of the IV All-Russian Scientific and Practical Conf. with International Participation „Information Systems and Technologies in Modeling and Control“. 2019. P. 193—202.
7. *Tatarnikova T. M., Poymanova E. D.* Differentiated capacity extension method for system of data storage with multilevel structure // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2020. Vol. 1, N 1. P. 66—73.
8. *Советов Б. Я., Татарникова Т. М., Пойманова Е. Д.* Организация многоуровневого хранения данных // Информационно-управляющие системы. 2019. № 2 (99). С. 68—75.
9. *Bogatyrev V. A.* Fault tolerance of clusters configurations with direct connection of storage devices // Automatic Control and Computer Sciences. 2011. Vol. 45, N 6. P. 330—337.
10. *Sovetov B. Ya., Tatarnikova T. M., Poymanova E. D.* Storage scaling management model // Information and Control Systems. 2020. N 5. P. 43—49.
11. *Poymanova E. D., Tatarnikova T. M.* Applying machine learning methods for forecasting // Wave Electronics and its Application in Information and Telecommunication Systems, WECONF 2020: Proc. of the Intern. Conf., St. Petersburg, 2020.
12. Свид. о рег. прогн. для ЭВМ 2019661945 RU. Программа прогноза наращивания емкости систем хранения данных / Е. Д. Пойманова, Т. М. Татарникова, Н. В. Яготинцева. 12.09.2019.

- Сведения об авторах**
- Екатерина Дмитриевна Пойманова** — канд. техн. наук; РГГМУ, кафедра информационных технологий и систем безопасности; доцент; E-mail: e.d.poymanova@gmail.com
- Татьяна Михайловна Татарникова** — д-р техн. наук, профессор; РГГМУ; Институт информационных систем и геотехнологий; директор; E-mail: tm-tatarn@yandex.ru
- Екатерина Витальевна Краева** — РГГМУ, кафедра информационных технологий и систем безопасности; ассистент; E-mail: kate.smitt.by@mail.ru

Поступила в редакцию
23.01.2021 г.

Ссылка для цитирования: Пойманова Е. Д., Татарникова Т. М., Краева Е. В. Модель управления хранением трафика данных // Изв. вузов. Приборостроение. 2021. Т. 64, № 5. С. 370—375.

MODEL OF DATA TRAFFIC STORAGE MANAGEMENT

E. D. Poymanova, T. M. Tatarnikova, E. V. Kraeva

Russian State Hydrometeorological University,
192007, St. Petersburg, Russia
E-mail: e.d.poymanova@gmail.com

A complex solution for managing traffic recording and storage in data storage systems is presented. A structure of a software and hardware complex for managing physical data storage for systems that can be used by owners of technological communication networks for storing traffic is proposed. Control mechanisms are considered, such as the data distribution across various media using Kohonen neural networks and forecasting the storage capacity increase using a statistical model and machine learning methods.

Keywords: traffic, data storage system, data distribution, physical data storage, machine learning, neural network, forecasting

REFERENCES

1. <https://tass.ru/ekonomika/9574021>. (in Russ.)
2. <https://digital.gov.ru/ru/pages/statistika-otrasli/>. (in Russ.)
3. <https://www.kommersant.ru/doc/4522028>. (in Russ.)
4. Farley M. *Building Storage Networks*, McGraw-Hill, Osborne, 2001.
5. *Information Storage and Management*, John Wiley & Sons Inc., 2016.
6. Tatarnikova T.M., Poymanova E.D. *CEUR Workshop Proceedings. ISTMC'2019 Selected Papers of the IV All-Russian Scientific and Practical Conference with International Participation "Information Systems and Technologies in Modeling and Control"*, 2019, pp. 193–202.
7. Tatarnikova T.M., Poymanova E.D. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, no. 1(1), pp. 66–73.
8. Sovetov B. Ya., Tatarnikova T.M., Poymanova E.D. *Information and Control Systems*, 2019, no. 2(99), pp. 68–75. (in Russ.)
9. Bogatyrev V.A. *Automatic Control and Computer Sciences*, 2011, no. 6(45), pp. 330–337.
10. Sovetov B.Ya., Tatarnikova T.M., Poymanova E.D. *Information and Control Systems*, 2020, no. 5, pp. 43–49.
11. Poymanova E.D., Tatarnikova T.M. *2020 Wave Electronics and its Application in Information and Telecommunication Systems*, WECONF 2020.
12. Certificate of computer program registration RU 2019661945, *Programma prognoza narashchivaniya yemkosti sistem khraneniya dannykh* (Program for Forecasting the Growth of Storage Capacity), E.D. Poymanova, T.M. Tatarnikova, N.V. Yagotintseva, 22.07.2019. (in Russ.)

Data on authors

- Ekaterina D. Poymanova** — PhD; Russian State Hydrometeorological University, Department of Information Technology and Security Systems; Associate Professor; E-mail: e.d.poymanova@gmail.com
- Tatiana M. Tatarnikova** — Dr. Sci., Professor; Russian State Hydrometeorological University, Institute of Information Systems and Geotechnologies; Head of the Institute; E-mail: tm-tatarn@yandex.ru
- Ekaterina V. Kraeva** — Russian State Hydrometeorological University, Department of Information Technology and Security Systems; Assistant; E-mail: kate.smitt.by@mail.ru

For citation: Poymanova E. D., Tatarnikova T. M., Kraeva E. V. Model of data traffic storage management. *Journal of Instrument Engineering*. 2021. Vol. 64, N 5. P. 370—375 (in Russian).

DOI: 10.17586/0021-3454-2021-64-5-370-375