

УДК 004.7

doi: 10.17586/2226-1494-2020-20-1-66-73

МЕТОДИКА ДИФФЕРЕНЦИРОВАННОГО НАРАЩИВАНИЯ ЕМКОСТИ СИСТЕМЫ ХРАНЕНИЯ ДАННЫХ С МНОГОУРОВНЕВОЙ СТРУКТУРОЙ

Т.М. Татарникова^{a,b}, Е.Д. Пойманова^c

^a Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург, 197376, Российская Федерация

^b Российский государственный гидрометеорологический университет, Санкт-Петербург, 195196, Российская Федерация

^c Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, 190000, Российская Федерация

Адрес для переписки: Tm-tatarn@yandex.ru

Информация о статье

Поступила в редакцию 04.11.19, принята к печати 01.12.19

Язык статьи — русский

Ссылка для цитирования: Татарникова Т.М., Пойманова Е.Д. Методика дифференцированного наращивания емкости системы хранения данных с многоуровневой структурой // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 1. № 1. С. 66–73. doi: 10.17586/2226-1494-2020-20-1-66-73

Аннотация

Предмет исследования. Предложена методика дифференцированного наращивания емкости хранилища данных, построенная на прогнозной модели временных рядов с оценкой объема поступающего на хранение трафика. Рассмотрено влияние структуры входящего потока данных на выбор модели прогноза. **Используемые подходы.** Система хранения данных формализована в виде матрицы, устанавливающей количество уровней хранения и количество носителей/томов на каждом уровне. Элементами матрицы являются метаданные файлов, которые хранятся на соответствующих носителях/томах многоуровневой системы хранения данных. Матрица визуализирует состояние хранилища данных в виде паттернов. Построение паттернов выполняется с помощью систематических срезов значений матрицы. Периодический анализ паттернов состояния хранилища данных позволяет оценить время достижения предельного значения емкости носителя. Прогнозная модель, положенная в основу методики дифференцированного наращивания емкости хранилища данных, учитывает структуру входящего потока данных. При наличии самоподобной структуры поступающего на хранение трафика реализуется прогнозная модель авторегрессии и проинтегрированного скользящего среднего. Для трафика без самоподобной структуры реализуется общая линейная модель прогноза временного ряда при известных прошлых значениях. Модель прогноза применяется отдельно для каждого носителя/тома уровня хранения. **Основные результаты.** Приведены особенности структуры трафика, поступающего на хранение. Проверены свойства самоподобия на примере LTE-трафика, демонстрирующие наличие распределений с «тяжелыми хвостами». С помощью модели авторегрессии и проинтегрированного скользящего среднего получены результаты прогноза объема поступающего на хранение трафика. Приведены прогнозные и реальные значения объема трафика, а также величина ошибки прогноза. Разработана методика дифференцированного наращивания емкости системы хранения данных, устанавливающая последовательность шагов анализа паттернов и структуры трафика, поступающего на хранение. **Практическая значимость.** Методика наращивания емкости хранилища данных учитывает многоуровневую организацию хранения и структуру поступающего потока данных, позволяет организовать дифференцированное наращивание емкости хранилища в соответствии с характеристиками файлов и обеспечением требований к времени гарантированного хранения.

Ключевые слова

многоуровневое хранение, система хранения данных, хранилище данных, структура трафика, паттерн состояния хранилища данных, модель прогноза, методика наращивания емкости хранилища

DIFFERENTIATED CAPACITY EXTENSION METHOD FOR SYSTEM OF DATA STORAGE WITH MULTILEVEL STRUCTURE

Т.М. Tatarnikova^{a,b}, Е.Д. Poymanova^c

^a Saint Petersburg Electrotechnical University “LETI”, Saint Petersburg, 197376, Russian Federation

^b Russian State Hydrometeorological University, Saint Petersburg, 195196, Russian Federation

^c Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, 190000, Russian Federation

Corresponding author: Tm-tatarn@yandex.ru

Article info

Received 04.11.19, accepted 01.12.19

Article in Russian

For citation: 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. 20, no. 1, pp. 66–73 (in Russian). doi: 10.17586/2226-1494-2020-20-1-66-73

Abstract

Subject of Research. The paper presents a method for differentiated capacity extension of the data warehouse. The method is built on a predictive model of time series with an estimate of volume for the traffic storage. The effect of the incoming data stream structure on the choice of the prediction model is considered. **Methods.** The storage system is presented in the form of a matrix specifying the number of storage levels and the number of carriers/volumes at each level. The matrix elements are metadata of the recorded files that are stored on the corresponding carriers/volumes of multilevel data storage system. The matrix visualizes the data storage state in the form of patterns. Patterning is performed by systematic slices of matrix values. Periodic analysis of the data warehouse state patterns gives the possibility to evaluate the time to reach the maximum value of the carrier capacity. The predictive model, which is the basis of the method for data warehouse differentiated capacity extension, takes into account the structure of the incoming data stream. In the presence of a self-similar structure of traffic for storage, a predictive model of auto-regression and an integrated moving average is implemented. For traffic without a self-similar structure, a general linear predictive model of the time series at known past values is implemented. The prediction model is applied separately for each storage carrier/volume. **Main Results.** Structure features of the traffic arriving for storage are given. Self-similarity properties are verified on the example of LTE-traffic, demonstrating the presence of “heavy-tailed” distributions. The prediction results for volume of traffic arriving for storage are obtained by the autoregressive model and the integrated moving average. The predictive and real values of the traffic volume are given, as well as the prediction error value. A technique for differentiated capacity extension of the data storage system is developed, which establishes a sequence of steps for analysis of patterns and the structure of traffic arriving for storage. **Practical Relevance.** The method for differentiated capacity extension of the data storage takes into account the multilevel organization of storage and the structure of the incoming data stream, which provides organizing a differentiated capacity extension in accordance with the characteristics of the files and ensuring the requirements for guaranteed storage time.

Keywords

multilevel storage, data storage system, data warehouse, traffic structure, data warehouse state pattern, prediction model, storage capacity extension method

Введение

В хранении данных нуждаются как отдельные пользователи, так и крупные организации, и государственные структуры. Это связано с ведением бизнеса, электронным документооборотом, работой аналитических систем [1].

Инфраструктуру хранения данных в виде IaaS-услуги (Infrastructure as a Service – инфраструктура как услуга) предлагают Dell EMC, Fujitsu, IBM. Решения этих компаний основаны на технологиях дубликации, зеркалирования, виртуализации, иерархического размещения файлов, призванных обеспечить гарантированное хранение. Инфраструктура хранения реализуется системами хранения данных (СХД), представляющими собой архитектуру подключения носителей данных различной физической природы. Совокупность носителей без детализации особенностей их подключения, реализации доступа и т. д. также называют хранилищем данных [2].

В архитектуре СХД определяют три уровня, каждый из которых предполагает свои технологии хранения: RAID (Redundant Array of Independent Disks — избыточ-

ный массив независимых дисков), автоматизированные библиотеки и носители длительного хранения (рис. 1) [3, 4]. СХД выделяют в отдельную подсистему вычислительного комплекса, например, центра обработки данных [5, 6].

При реализации процесса хранения требуется своевременное выделение необходимой емкости для размещения данных, причем для каждого уровня, т. е. дифференцированно. Эта задача решается функциями управления СХД, в частности своевременным наращиванием емкости хранилища данных [7, 8].

Постановка задачи

В работе предлагается методика дифференцированного наращивания емкости хранилища данных, построенная на прогнозной модели временных рядов с оценкой объема трафика, поступающего на вход СХД. Прогноз позволяет своевременно увеличивать емкость СХД и предотвращать потерю данных, поступающих на хранение [9, 10].

Систему хранения данных S формально представим как множество параметров $\{C, B, V\}$, описывающих:



F — поток файлов, требующих хранения

RAID	Том 1	Том 2	...	Том n
Автоматизированные библиотеки	Носитель 1	Носитель 2	...	Носитель n
Носители длительного хранения	Носитель 1	Носитель 2	...	Носитель n

Рис. 1. Многоуровневая структура хранилища данных

S — структуру СХД; \mathbf{B} — состояние СХД; \mathbf{V} — характеристику емкости носителей/томов СХД, в байтах.

Пусть любой файл, поступающий на хранение, характеризуется множеством параметров $\{t, f, \lambda\}$, где t — требуемое время хранения, в годах — задается автоматически типом файла type; f — размер файла, в байтах; λ — частота обращения к файлам, запросов/час.

Тогда поток файлов, поступающий в СХД на хранение, обозначим как $F = \{t, f, \lambda\}$.

Представим структуру СХД в виде матрицы \mathbf{R} размером $m \times n$. Элементами матрицы \mathbf{R} являются метаданные $\{t, f, \lambda\}$ файлов, которые хранятся на соответствующих носителях/томах многоуровневой СХД. В свою очередь каждая ячейка \mathbf{R} имеет следующие характеристики:

- максимальное значение емкости V_{\max} , которое соответствует реальной емкости используемого носителя (тома носителя);

- предельное значение емкости V_{\lim} ($V_{\lim} < V_{\max}$), при достижении которого необходимо производить наращивание емкости хранилища данных.

Кроме того, зададим пограничные значения частоты обращения к файлам $\lambda_1, \dots, \lambda_{m-1}$, при преодолении которых осуществляется миграция файлов по уровням системы.

Предлагаемая в работе модель прогноза основана на анализе срезов состояния хранилища данных и представляет собой паттерн поведения СХД, исходя из прогноза поведения ячеек матрицы \mathbf{R} [11, 12].

Для актуализации паттерна поведения СХД необходим периодический анализ метаданных о размере файлов каждой ячейки матрицы хранения для оценивания ее текущей емкости V_{cur} [13]. С течением времени состояние ячеек матрицы меняется. Это связано, во-первых, с входящим потоком файлов в СХД, а во-вторых, с миграцией файлов внутри физического хранилища. Таким образом, задача прогнозирования заключается в нахождении времени достижения t_{\lim} предельной емкости V_{\lim} и времени достижения t_{\max} максимальной емкости V_{\max} каждой ячейки матрицы \mathbf{R} .

Значение V_{\lim} и V_{\max} , заданные в паттерне состояний матрицы \mathbf{R} , математически можно выразить следующим образом:

$$V_{\lim} = \int_1^{t_{\lim}} f(t) dt, \quad (1)$$

$$V_{\max} = \int_1^{t_{\max}} f(t) dt, \quad (2)$$

где $f(t)$ — функция входящего потока данных.

Функция $f(t)$ математически может быть не определена, поэтому предлагается выразить значения V_{\lim}

и V_{\max} исходя из метода правых прямоугольников. Тогда

$$V_{\lim} = \int_1^{t_{\lim}} f(t) dt = h \sum_1^{t_{\lim}} f(t), \quad (3)$$

$$V_{\max} = \int_1^{t_{\max}} f(t) dt = h \sum_1^{t_{\max}} f(t), \quad (4)$$

где h — шаг разбиения, равный единице минимального выбранного масштаба времени.

Поскольку функция $f(t)$ не имеет первообразной, то предлагается вычислять значения t_{\lim} и t_{\max} программно, методом подстановки, т. е. искать такие t_{\lim} и t_{\max} , при которых соответствующие V_{\lim} и V_{\max} достигнут установленное значение. Данные расчеты производятся для каждой ячейки матрицы \mathbf{R} .

В результате паттерн СХД (3×3) будет выглядеть следующим образом (рис. 2).

Структура трафика, поступающего на хранение

При построении модели прогноза наращивания СХД необходимо рассматривать структуру входящего потока данных, который является гетерогенным и представляет собой смесь из речевых, текстовых и мультимедиа данных. Рассматривая входящий поток данных на некотором отрезке времени, можно проследить наличие пульсаций – смену периодов поступления большего или меньшего потока данных на вход системы. Пульсации получаются в результате неравномерной активности пользователей СХД, связанной с рабочими часами, выходными, праздниками, периодами отпусков или другими событиями. В случае, когда пульсации представляют некую фрактальную (самоподобную) структуру, есть возможность составления длительного прогноза входящего потока данных в разном масштабе времени (рис. 3) [14].

$type_1,$ $V_{\max11}, V_{\lim11},$ $t_{\max11}, t_{\lim11},$ $V_{\text{cur}11}$	$type_1,$ $V_{\max12}, V_{\lim12},$ $t_{\max12}, t_{\lim12},$ $V_{\text{cur}12}$	$type_1,$ $V_{\max13}, V_{\lim13},$ $t_{\max13}, t_{\lim13},$ $V_{\text{cur}13}$
$type_2,$ $V_{\max21}, V_{\lim21},$ $t_{\max21}, t_{\lim21},$ $V_{\text{cur}21}$	$type_2,$ $V_{\max22}, V_{\lim22},$ $t_{\max22}, t_{\lim22},$ $V_{\text{cur}22}$	$type_2,$ $V_{\max23}, V_{\lim23},$ $t_{\max23}, t_{\lim23},$ $V_{\text{cur}23}$
$type_3,$ $V_{\max31}, V_{\lim31},$ $t_{\max31}, t_{\lim31},$ $V_{\text{cur}31}$	$type_3,$ $V_{\max32}, V_{\lim32},$ $t_{\max32}, t_{\lim32},$ $V_{\text{cur}32}$	$type_3,$ $V_{\max33}, V_{\lim33},$ $t_{\max33}, t_{\lim33},$ $V_{\text{cur}33}$

Рис. 2. Паттерн системы хранения данных

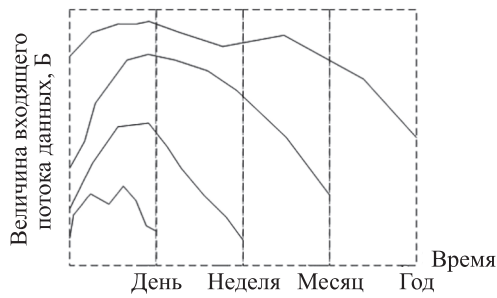


Рис. 3. Фрактальная (самоподобная) структура входящего потока данных

Методика прогнозирования наращивания емкости СХД на основе анализа потока входящих данных и их миграции

Предлагаемая методика прогнозирования наращивания емкости основывается на паттерне поведения СХД и анализе срезов ее состояния и состоит из следующих этапов:

Анализ текущего состояния системы.

1.1. Построение нулевого паттерна СХД, включающего определение:

- размера матрицы \mathbf{R} ;
- максимального размера емкости ячеек V_{\max} ;
- предельного значения емкости ячеек $V_{\lim} < V_{\max}$, при достижении которого выполняется наращивание;
- пограничных значений частоты обращения к файлам $\lambda_1, \dots, \lambda_{m-1}$, при преодолении которых будет осуществляться миграция файлов по уровням СХД (строкам матрицы \mathbf{R}).

1.2. Построение актуального паттерна СХД, включающего анализ метаданных:

- о размерах файлов V_{cur} , записанных в ячейках матрицы \mathbf{R} ;
- о типах файлов *type*, задающих предполагаемое время хранения.

2. Выявление статистических свойств входящего потока данных.

2.1. Статистическая оценка признаков самоподобия поступающего в СХД потока файлов:

- медленно затухающей дисперсии;
- долговременной зависимости;
- наличия распределения с тяжелым «хвостом», которым характеризуется распределением промежутков времени между соседними поступлениями файлов в СХД.

Анализ проводится на основе изменения метаданных файлов в разном масштабе времени: час, день, неделя, месяц, год. В идеальном случае предполагается, что ведется соответствующая статистика.

2.2. Визуальная оценка структуры входящего потока данных:

2.2.1. Построение для каждой ячейки матрицы \mathbf{R} графиков $F(t)$ в разном масштабе времени, где F — это количество входящего потока, измеряемого в мегабайтах, поступающего на вход ячейки матрицы хранения, а t — время, за которое проводятся наблюдения [12];

2.2.2. Первичная (визуальная) оценка полученных графиков $F(t)$. Если на разных шкалах времени струк-

тура графиков $F(t)$ повторяется (наличие самоподобности), то переход на п. 2.3, иначе переход на п. 3.1.

2.3. Определение параметра формы распределения входящего потока данных:

2.3.1. Построение графика функции распределения вероятностей входящего потока $F(x) = 1 - F(x) = P(X < x)$ в логарифмическом масштабе;

2.3.2. Нахождение параметра формы распределения α решением уравнения регрессии. Если $\alpha \in [0; 2]$, то распределение входящего потока данных имеет свойство «тяжелого хвоста», переход на п. 2.3.3, иначе на п. 3.1;

2.3.3. Оценка показателя Херста по формуле:

$$H = \frac{3 - \alpha}{2}. \quad (5)$$

Если $H \in [0,5; 1]$, то входящий поток данных — самоподобный, переход на п. 2.3.4, иначе на п. 3.1;

2.3.4. Увеличение масштаба времени с применением значения H , при необходимости повторить п. 2.2.2.

Построение паттерна поведения системы (прогнозной модели).

3.1. Реализация общей линейной модели прогноза временного ряда при известных прошлых значениях для графика без самоподобной структуры:

$$X_t = X_t^* + \varepsilon_t,$$

где X_t^* — линейный предиктор; ε_t — случайная ошибка линейного предиктора.

3.2. Для графика с самоподобной структурой реализация модель авторегрессии и проинтегрированного скользящего среднего ARIMA(p, d, q):

$$\varphi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t,$$

где $\varphi(\cdot)$, $\theta(\cdot)$ — полиномы степени p и q ; B — лаговый оператор, $B^j X_t = X_{t-j}$, $B^j \varepsilon_{t-j}$, $j = 0, \pm 1, \dots$; d — порядок взятия последовательности разности

$$\Delta X_t = X_{t-1} - X_t = (1 - B)X_t,$$

$$\Delta^2 X_t = \Delta^2 X_{t+1} - \Delta X_t = (1 - B)^2 X_t, \dots$$

Для выбора параметров p, d, q используется автокорреляционный анализ: d выбирается исходя из количества применяемых операций последовательной разности (как правило, $d = 2$, поскольку взятие вторых разностей позволяет привести любой нестационарный ряд к стационарному виду), p выбирается из модели частичной автокорреляции, которая позволяет измерить связь между текущим уровнем временного ряда и его предыдущими значениями, q выбирается из модели автокорреляции, которая позволяет измерить зависимость между значением функции и ее сдвинутой копией от величины временного сдвига.

Расчет времени преодоления пограничных значений емкости ячеек матрицы \mathbf{R} .

4.1. Нахождение времени достижения предельной емкости t_{\lim} и времени достижения максимальной емкости t_{\max} каждой ячейки в соответствии с выражениями (1)–(4).

Мониторинг состояния СХД и коррекция модели прогноза.

5.1. Построение срезов состояний системы. Если срезы состояния системы близки к паттерну поведения, то осуществляется переход на п. 4, в этом случае прогноз верен и корректировка прогнозной модели не требуется, иначе переход на п. 5.2.

5.2. Корректировка прогнозной модели на основании новых статистических данных в случае существенных отклонений от паттерна поведения. Переход на п. 1.

Детализация некоторых шагов методики прогнозирования емкости хранилища при поступлении самоподобного трафика

В эксперименте участвовали данные LTE-трафика, полученные от компании МТС (Мобильные Теле-Системы) г. Санкт-Петербург в период с 08.08.2018 г. по 13.08.2018 г. (рис. 4).

Визуальная оценка структуры входящего потока данных. Примеры визуальной оценки структуры входящего потока суточных данных представлены на рис. 5.

Определение параметра формы распределения входящего потока данных. График функции распределения вероятностей входящего потока в логарифмическом масштабе приведен на рис. 6.

В примере на рис. 6, *a* полученное уравнение регрессии показывает, что значение параметра α находится в интервале $[0; 2]$, следовательно, распределение входящего потока данных имеет свойство «тяжелого хвоста». Согласно (5) показатель Херста $H = 0,855$,

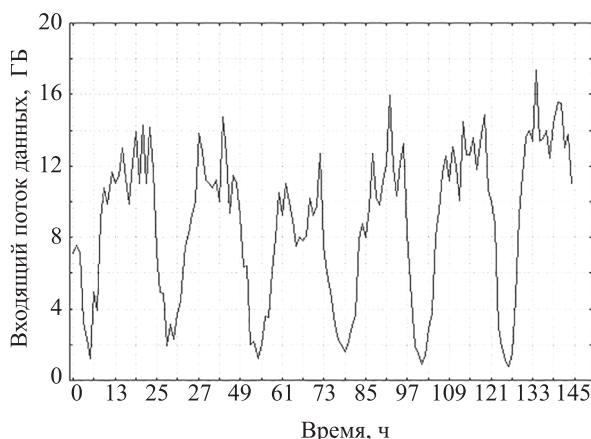


Рис. 4. Входящий поток на основе данных LTE-трафика компании МТС

следовательно, входящий поток данных имеет свойства самоподобия.

В табл. 1 показаны результаты расчета дисперсии результатов оценки D , среднего значения M и коэффициента корреляции r для начального и агрегированного трафика. Приведенные результаты демонстрируют свойства распределений с «тяжелыми хвостами».

Определение параметров p, d, q . Поскольку операция последовательной разницы была применена один раз, то $d = 1$. Параметр p выбирается из модели частичной автокорреляции и, в рассматриваемом при-

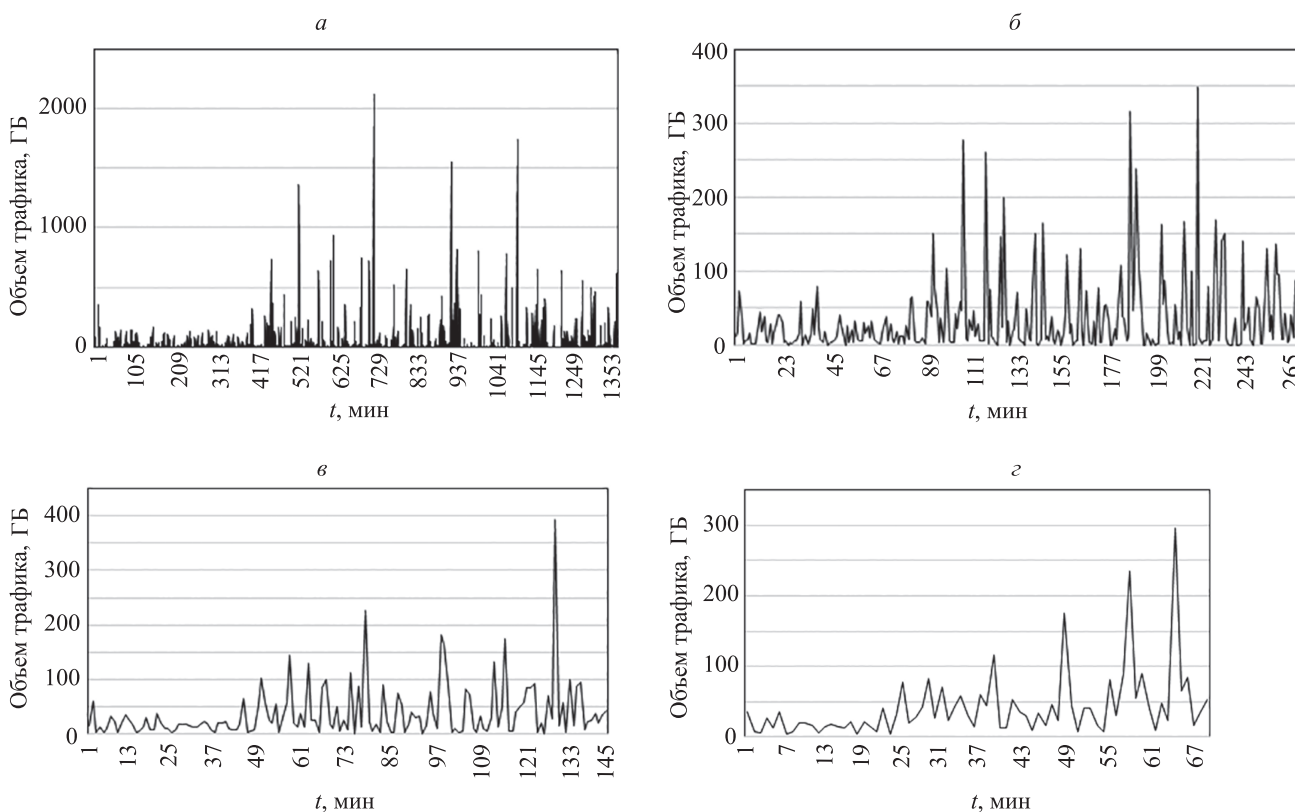


Рис. 5. Примеры графиков объемов входящего потока в разном масштабе времени в течение: 1440 мин (*a*); 276 мин (*б*); 145 мин (*в*); 70 мин (*г*)

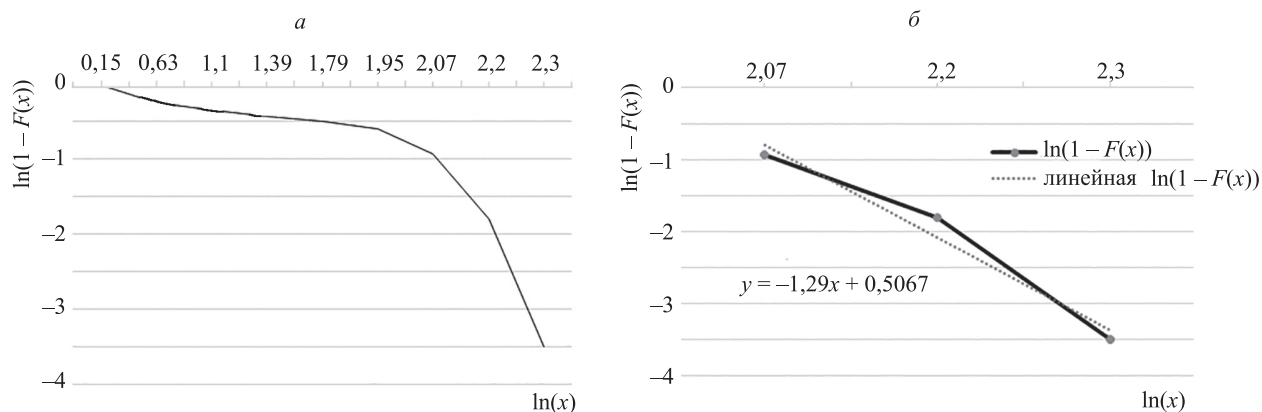


Рис. 6. Пример графика функции распределения вероятностей: обычный вид (а); «хвост» функции распределения (б)

Таблица 1. Результаты расчетов показателей трафика

Входящий поток данных	$D(X)$	$M(X)$	$r(k), k = 2$
Исходный поток данных	4149,47	37,82	0,0233
Агрегированный входящий поток за 5 мин	4341,17	37,25	0,0712
Агрегированный входящий поток за 10 мин	3354,93	40,39	0,0678
Агрегированный входящий поток за 20 мин	3991,34	38,21	0,0349

мере, $p = 1$ (первая значимая величина ряда функции). Параметр q выбирается аналогично из модели автокорреляции и $q = 1$.

Результат прогноза на один период, полученный с помощью модели ARIMA(1,1,1), приведен на рис. 7.

В табл. 2 приведены прогнозные и реальные значения объема трафика, а также величина ошибки прогноза как абсолютная разница между прогнозными и реальными значениями объема трафика.

Приведенный эксперимент является примером построения модели ARIMA, иллюстрирующим порядок построения прогнозной модели.



Рис. 7. Модель прогноза входящего потока данных

Таблица 2. Данные модели прогноза модели ARIMA(1,1,1)

Номер значения временного ряда	Прогнозный трафик	Исходный трафик	Разность	Ошибка, %
120	14,93372	10,66000	4,27372	29
121	11,21004	10,00000	1,21004	11
122	8,61671	8,72000	0,10329	1
123	7,51864	6,97000	0,54864	7
124	5,44676	4,87000	0,57676	11
125	5,34649	4,18000	0,16649	3
126	5,08029	4,80000	0,28029	5
127	6,19234	5,45000	0,74234	12
128	7,03461	6,57000	0,46461	7
129	9,56586	9,19000	0,37586	4

Таблица 2. Продолжение

130	10,89524	11,51000	0,61476	6
131	12,74505	13,61000	0,86495	7
132	13,98409	14,00000	0,01591	0
133	13,82446	13,40000	0,42446	3
134	14,92202	17,36000	2,43798	16
135	14,43590	13,38000	1,05590	7
136	13,09138	13,50000	0,40862	3
137	14,23317	14,00000	0,23317	2
138	14,01566	12,45000	1,56566	11
139	13,94582	14,36000	0,41418	3
140	15,60675	15,58000	0,02675	0
141	14,63978	15,50000	0,86022	6
142	14,14796	13,00000	1,14796	8
143	15,48867	13,80000	1,68867	11

Заключение

Структура системы хранения данных представлена в виде математической матрицы, число ячеек которой соответствует количеству уровней иерархического хранения и числу носителей уровня. Показано, что задача прогнозирования заключается в нахождении времени достижения предельной емкости и времени достижения максимальной емкости каждой ячейки матрицы.

В зависимости от структуры трафика, поступающего в хранилище данных, выбраны две модели прогнозирования: общая линейная модель для трафика без самоподобной структуры и модель авторегрессии

и проинтегрированного скользящего среднего для самоподобного трафика.

Планирование дифференцированного наращивания емкости хранилища основано на учете паттернов состояния системы хранения данных. Для построения паттернов рекомендуется проводить мониторинг состояния системы хранения данных с помощью систематических срезов значений матрицы.

Предложенная методика прогнозирования наращивания емкости системы хранения данных необходима для своевременного выделения пространственного ресурса и снижения потерь поступающего на хранение трафика.

Литература

1. Проскуряков Н.Е., Ануфриева А.Ю. Анализ и перспективы современных систем хранения цифровых данных // Известия Тульского государственного университета. Технические науки. 2013. № 3. С. 368–377.
2. Information Storage and Management. 2nd ed. New Jersey: John Wiley & Sons Inc., 2016. 544 p.
3. Farley M. Building Storage Networks. 2nd ed. Osborne, McGraw-Hall, 2001. 576 p.
4. Леонов В. Google Docs, Windows Live и другие облачные технологии. М.: Эксмо-Пресс, 2012. 304 с.
5. Богатырев В.А., Богатырев С.В., Богатырев А.В. Надежность кластерных вычислительных систем с дублированными связями серверов и устройств хранения // Информационные технологии. 2013. № 2. С. 27–32.
6. Mesnier M., Ganger G., Riedel E. Object-based storage // IEEE Communications Magazine. 2003. V. 41. N 8. P. 84–90. doi: 10.1109/MCOM.2003.1222722
7. Carr N.G. The Big Switch: Our New Digital Destiny. WW Norton & Company, 2008. 258 p.
8. Бурмистров В.Д., Заковряшин Е.М. Создание хранилища данных для распределенной системы // Молодой ученый. 2016. № 12. С. 143–147.
9. Buyya R., Broberg J., Goscinski A.M. Cloud Computing: Principles and Paradigms. New Jersey: John Wiley & Sons Inc., 2011. 637 p.
10. Советов Б.Я., Татарникова Т.М., Пойманова Е.Д. Организация многоуровневого хранения данных // Информационно-управляющие системы. 2019. № 2. С. 68–75. doi: 10.31799/1684-8853-2019-2-68-75

References

1. Proskuryakov N.E., Anufrieva A.Y. Analysis and prospects of modern systems of storage of figures. *News of the Tula state university. Technical sciences*, 2013, no. 3, pp. 368–377. (in Russian)
2. *Information Storage and Management*. 2nd ed. New Jersey, John Wiley & Sons Inc., 2016, 544 p.
3. Farley M. *Building Storage Networks*. 2nd ed. Osborne, McGraw-Hall, 2001, 576 p.
4. Leonov V. *Google Docs, Windows Live, and other cloud technologies*. Moscow, EKSMO Publ., 2012, 304 p. (in Russian)
5. Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Reliability clusters computing systems with the duplicated communications of servers and storage devices. *Information Technology*, 2013, no. 2, pp. 27–32. (in Russian)
6. Mesnier M., Ganger G., Riedel E. Object-based storage. *IEEE Communications Magazine*, 2003, vol. 41 no. 8, pp. 84–90. doi: 10.1109/MCOM.2003.1222722
7. Carr N.G. *The Big Switch: Our New Digital Destiny*. WW Norton & Company, 2008, 258 p.
8. Burmistrov V.D., Zakovryashin E.M. Creating a data warehouse for a distributed system. *Molodoi Uchenyi*, 2016, no. 12, pp. 143–147. (in Russian)
9. Buyya R., Broberg J., Goscinski A.M. *Cloud Computing: Principles and Paradigms*. New Jersey, John Wiley & Sons Inc., 2011, 637 p.
10. Sovetov Ya.B., Tatarnikova T.M., Poymanova E.D. Organization of multi-level data storage. *Informatsionno-Upravliaiushchie Sistemy*, 2019, no. 2, pp. 68–75. (in Russian). doi: 10.31799/1684-8853-2019-2-68-75

11. Kish L.B., Granqvist C.G. Does information have mass? // *Proceedings of the IEEE*. 2013. V. 101. N 9. P. 1895–1899. doi: 10.1109/JPROC.2013.2273720
12. Morville P., Callender J. *Search Patterns: Design for Discovery*. O'Reilly Publ., 2010. 192 p.
13. Stacey M., Salvatore J., Jorgensen A. *Visual Intelligence: Microsoft Tools and Techniques for Visualizing Data*. New Jersey: John Wiley & Sons Inc., 2013. 432 p.
14. Poymanova E.D., Tatarnikova T.M. Models and methods for studying network traffic // *2018 Wave Electronics and its Application in Information and Telecommunication Systems, WECONF 2018*. 2018. P. 8604470. doi: 10.1109/WECONF.2018.8604470
11. Kish L.B., Granqvist C.G. Does information have mass? *Proceedings of the IEEE*, 2013, vol. 101, no. 9, pp. 1895–1899. doi: 10.1109/JPROC.2013.2273720
12. Morville P., Callender J. *Search Patterns: Design for Discovery*. O'Reilly Media, 2010, 192 p.
13. Stacey M., Salvatore J., Jorgensen A. *Visual Intelligence: Microsoft Tools and Techniques for Visualizing Data*. New Jersey, John Wiley & Sons Inc., 2013, 432 p.
14. Poymanova E.D., Tatarnikova T.M. Models and methods for studying network traffic. *2018 Wave Electronics and its Application in Information and Telecommunication Systems, WECONF 2018*, 2018, pp. 8604470. doi: 10.1109/WECONF.2018.8604470

Авторы

Татарникова Татьяна Михайловна — доктор технических наук, доцент, профессор, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург, 197376, Российская Федерация; заведующий кафедрой, Российский государственный гидрометеорологический университет, Санкт-Петербург, 195196, Российская Федерация, Scopus ID: 36715607400, ORCID ID: 0000-0002-6419-0072, Tm-tatarn@yandex.ru
Пойманова Екатерина Дмитриевна — старший преподаватель, Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, 190000, Российская Федерация, ORCID ID: 0000-0002-7903-2480, e.d.poymanova@gmail.com

Authors

Tatiana M. Tatarnikova — D.Sc., Associate Professor, Professor, Saint Petersburg Electrotechnical University “LETI”, Saint Petersburg, 197376, Russian Federation; Head of Chair, Russian State Hydrometeorological University, Saint Petersburg, 195196, Russian Federation, Scopus ID: 36715607400, ORCID ID: 0000-0002-6419-0072, Tm-tatarn@yandex.ru

Ekaterina D. Poymanova — Senior Lecturer, Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, 190000, Russian Federation, ORCID ID: 0000-0002-7903-2480, e.d.poymanova@gmail.com