

Н. Ю. Братченко [N.Yu. Bratchenko],
В. П. Мочалов [V. P. Mochalov],
И. С. Палканов [I. S. Palkanov]

УДК 621-395.4

DOI: 10.37493/2307-910X.2022.2.3

**МЕТОД БАЛАНСИРОВКИ НАГРУЗКИ
ВЫЧИСЛИТЕЛЬНОГО КЛАСТЕРА
ЦЕНТРА ОБРАБОТКИ ДАННЫХ**

**METHOD OF LOAD BALANCING FOR
COMPUTER CLUSTER OF DATA
PROCESSING CENTER**

Северо-Кавказский федеральный университет, Ставрополь, Россия
North Caucasus Federal University, Pyatigorsk, Russia, e-mail: n.b.20062@yandex.ru ru.

В статье представлено описание метода балансировки нагрузки вычислительного кластера центра обработки данных (ЦОД) в основу которого положен вероятностный подход упреждающего прогнозирования состояний пакетного трафика, сформированный на основе результатов его статистического, нелинейного и спектрального анализа. Фрактальные свойства сетевого трафика являются обоснованием возможности предсказания, позволяют с достаточно большой вероятностью прогнозировать появление на отдельных временных интервалах всплесков и спадов его активности, выявление периодов возможной перегрузки серверов и сетевого оборудования и делают возможным разработку методов эффективного планирования и распределения задач внутри ЦОД, обеспечение статистически равномерной загрузки его функциональных элементов. Спектральный анализ временного ряда проводится по нормированным отклонениям фактических уровней от сглаженных. Отсутствие существенных пиков спектральных оценок говорит об отсутствии периодических колебаний. Показано, что суммирование циклов разного периода динамики временного ряда, основанное на использовании наиболее значимых гармоник спектра, определяет моменты возникновения последующих аномалий его развития. В основу процесса выявления существенных гармоник спектра положено исследование его спектральной плотности мощности с помощью преобразования Фурье. Разработанный метод способен обеспечить решение задачи эффективного планирования и распределения задач вычислительного кластера ЦОД с целью оптимизации использования ресурсов, ускорения времени выполнения задач и сокращения расходов на обработку приложений.

Ключевые слова: пакетный трафик, временные ряды, фракталы, балансировка нагрузки, функция автокорреляции, гармонический анализ, нелинейная динамика.

The article presents a description of the load balancing method for a computing cluster of a data processing center (DPC), which is based on a probabilistic approach to proactive forecasting of packet traffic states, formed on the basis of the results of its statistical, nonlinear and spectral analysis. The fractal properties of network traffic are the rationale for the possibility of prediction, allow with a fairly high probability to predict the appearance of bursts and drops in its activity at certain time intervals, identify periods of possible overload of servers and network equipment, and make it possible to develop methods for effective planning and distribution of tasks within the data center, ensuring a statistically uniform loading its functional elements. The spectral analysis of the time series is carried out according to the normalized deviations of the actual levels from the smoothed ones. The absence of significant peaks in the spectral estimates indicates the absence of periodic fluctuations. It is shown that the summation of cycles of different periods of the dynamics of the time series, based on the use of the most significant harmonics of the spectrum, determines

the moments of occurrence of subsequent anomalies in its development. The process of identifying significant harmonics of the spectrum is based on the study of its spectral power density using the Fourier transform. The developed method is able to provide a solution to the problem of efficient planning and distribution of tasks in a data center computing cluster in order to optimize the use of resources, speed up task execution time and reduce application processing costs.

Keywords: packet traffic, time series, fractals, load balancing, autocorrelation function, harmonic analysis, non-linear dynamics.

Введение

Вычислительные ресурсы ЦОД облачных систем реализуются в виде кластеров серверов и системы распределения и балансировки нагрузки. Задача системы распределения и балансировки нагрузки состоит в реализации метода, обеспечивающего примерно равную вычислительную загрузку элементов информационной системы ЦОД, а также минимальных затрат на передачу данных. Целью данного исследования является повышение эффективности функционирования ЦОД, предоставляющих информационные услуги, за счет применения метода управления пакетным трафиком на основе его фрактального и гармонического анализа. Структурная схема информационного кластера ЦОД, представленная на рисунке 1, содержит множество серверов и систему балансировки нагрузки, осуществляющую распределение запросов на основе информации мониторинга состояний серверов.

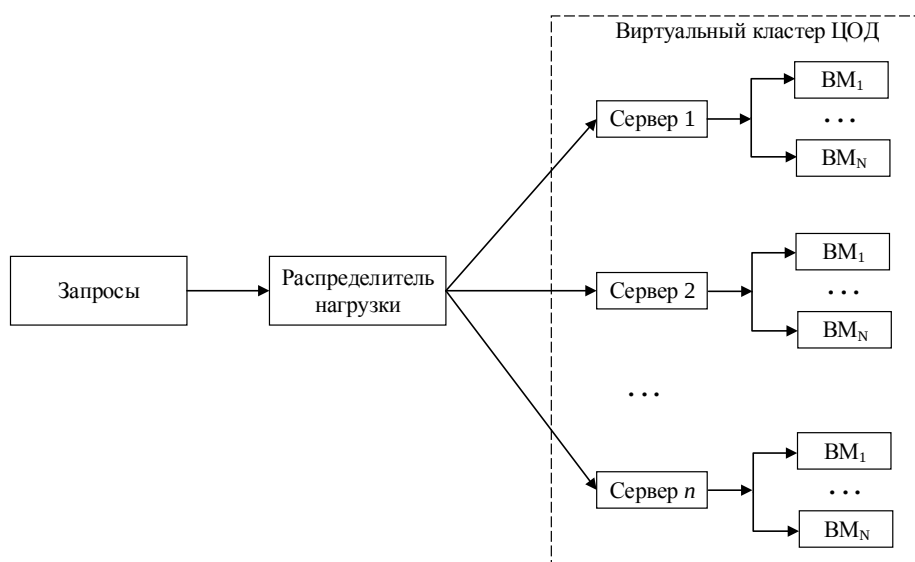


Рисунок 1. Структурная схема информационного кластера ЦОД/ Fig. 1. Structural diagram of the data center information cluster

Одной из важнейших причин, влияющих на эффективность работы ЦОД, являются аномалии сетевого трафика, заключающиеся в его фрактальности (самоподобии), частых всплесках и спадах активности, циклической составляющей, мощных пиковых выбросах и, в результате, перегрузке системы. Обнаружение подобных аномалий, своевременный прогноз времени их появления в будущем, для принятия мер по обеспечению качества обслуживания, вызывает необходимость создания более эффективных методов управления распределенной системой кластеров ЦОД. Известно [1,3,5], что свойства масштабной инвариантности сетевого трафика способны обеспечить, путем анализа на небольшом отрезке времени, прогноз его поведения на более длительных интервалах, обеспечивая этим реализацию эффективного планирования и распределения задач внутри ЦОД, статистически равномерную загрузки его функциональных элементов. Мерой длительности долговременной зависимости случайного процесса, обеспечивающей определение его фрактальности (самоподобия), наличие циклов, долговременной памяти, устойчивости является параметр самоподобия Херста H . Кратко-

временный прогноз состояний сетевого трафика, обеспечивают также преобразование Фурье и спектральный анализ [2,4]. Основное преимущество спектрального анализа заключается в его способности выявлять наиболее значимые гармоники спектра исследуемого процесса. Выделение наиболее существенных гармоник временного ряда, характеризующих динамику сетевого трафика, основано на оценке спектральной плотности мощности процесса с помощью дискретного преобразования Фурье, получения комплексных амплитуд ряда данных и последующего вычисления его спектра мощности. Это разложение представляет собой спектр динамики сетевого трафика. Точки максимума значений амплитуд спектра указывают на циклы периодических колебаний различной длины. Если выделить данные циклы из общего спектра сетевого трафика, а затем просуммировать их, то можно определить дальнейшую динамику развития аномалий трафика, время возникновения последующих мощных всплесков его активности, предсказать параметры наиболее вероятной пиковой нагрузки и моменты ее появления. В данном исследовании предлагается метод распределения и балансировки нагрузки, обеспечивающий повышение эффективности функционирования ЦОД, путем решения задачи кратковременного прогноза состояний сетевого трафика во времени и формирования на этой основе управляющих решений, направленных на поддержание равномерной загрузки оборудования ЦОД.

Методология исследования

Экспериментальные исследования [6,9,12] подтверждают факт фрактальной (самоподобной) структуры сетевого трафика. Известно, что фрактальные сетевые процессы обладают долговременной зависимостью, что выражается в практически бесконечном интервале корреляции и обеспечивает возможность прогноза его состояния на последующих временных интервалах [14]. Для реализации системы эффективной балансировки нагрузки, обеспечения статистически равномерной загрузки множества серверов кластеров ЦОД, необходимо учитывать структуру и свойства сетевого трафика, прогнозировать величину возможных скачков интенсивности нагрузки. Решение данной задачи возможно путем применения динамически изменяющегося алгоритма распределения и балансировки нагрузки, построенного на основе статистического анализа входящего в систему сетевого трафика, оценке степени его фрактальности, рядов Фурье и гармонического анализа. Данные мониторинга входной нагрузки ЦОД, используемые для реализации управляющего воздействия на систему балансировки нагрузки, могут быть представлены как числовые ряды, характеризующие изменения ее параметров во времени. Спектральный анализ позволяет из всех M гармоник спектра полученного временного ряда определить наиболее значимые, находящиеся в различных его выборках и имеющие самую большую амплитуду. Периоды наиболее значимых гармоник спектра определяют периодичность цикла. Если выделить данные гармонические составляющие спектра из его общей динамики, а затем представить их в виде суммы циклов разного периода, то можно определить дальнейшую динамику развития процесса, а также обеспечить возможность прогноза его дальнейшего развития. Алгоритм решения данной задачи включает в себя следующие шаги:

1. Мониторинг входной нагрузки ЦОД, представление ее в виде временного ряда, определение с помощью показателя Херста степени фрактальности пакетного трафика.
2. Определение с помощью метода наименьших квадратов линейного тренда временного ряда.
3. Удаление тренда из исходного временного ряда.
4. Из вновь сформированного временного ряда выделение гармоник, соответствующих наибольшему коэффициенту детерминации и добавление их в модель динамики временного ряда.
5. Расчет статистики Дарбина-Уотсона на наличие и уровень автокорреляции. Подтверждение допустимого качества модели регрессии.
6. Определение циклов, прогноз их развития.

7. На основе проведенного анализа фрактальных свойств сетевого трафика, а также данных о загруженности серверов, осуществление нового распределения нагрузки.

Материалы и методы исследования

Исследование показателя Херста.

Процесс поступления сетевого трафика в систему распределения и балансировки нагрузки ЦОД описывается моделью [7,8,10]

$$Y_t = F_{TP}(t) + \sum_{k=1}^M Y_k + \varepsilon_t \quad (1)$$

где $F_{TP}(t)$ – долговременная тенденция, тренд;

Y_k – гармоники ряда Фурье;

ε_t – случайная величина.

Присутствие тренда на интервале агрегации пакетного трафика, представленного в виде временного ряда, определяет H -показатель Херста, являющейся характеристикой устойчивости статистического процесса, оценкой корреляционной связи между его элементами и определяемый путем R/S анализа. Близость параметра H к 1 определяет достаточную трендоустойчивость процесса и возможность прогноза степени его изменения во времени. Основной формулой R/S анализа является выражение [11]

$$R/S = (a \cdot N)^H, \quad (2)$$

где H – показатель Херста;

N – выборка длины N ;

S – стандартное отклонение полученных измерений;

R – размах отношений $R = \max(Z_u) - \min(Z_u)$;

Z_u – накопленное отклонение ряда от среднего x_{cp} ;

a – константа;

$$S = \sqrt{N^{-1} \cdot \sum_{i=1}^N (x_i - x)^2}.$$

Логарифмируя полученное выражение, получаем

$$H = \log(R/S) / \log(N/2).$$

При $0,5 < H < 1$ имеем персистентный или трендоустойчивый ряд. Влияние настоящего на будущее описывается мерой корреляции [13]

$$C = 2^{2H-1} - 1.$$

Коэффициент Херста H является мерой длительности долговременной зависимости и описывает все остальные фрактальные параметры исследуемого процесса [15,17]:

- фрактальная размерность $D = 2 - H$;
- корреляционный параметр $\beta = 2(1 - H)$;
- спектральный показатель $b = 2H + 1$;
- фрактальный показатель $a = 3 - 2H$.

У самоподобного процесса $x(t)$ с $0,5 < H < 1$ корреляционная функция гиперболически затухает [12]

$$R(K) = \frac{\sigma^2}{2} \left[(K+1)^{2H} - 2K^{2H} + (K-1)^{2H} \right].$$

По определению коэффициент корреляции

$$\tau(K) = R(K) / R(\sigma) = R(K) / \sigma^2,$$

отсюда автокорреляционная функция (АКФ) будет иметь вид

$$\tau(K) = \frac{1}{2} \left[(K+1)^{2H} - 2K^{2H} + (K-1)^{2H} \right]. \quad (3)$$

Расчет АКФ необходимо осуществлять с целью оценки характера убывания зависимости элементов временного ряда.

Численные значения АКФ могут быть получены по формуле [14]

$$\tau(K) = \frac{\sum_{i=1}^{N-K} (x_i - \bar{x})(x_{i+K} - \bar{x})}{(N-K)\sigma^2(x)}, \quad (4)$$

где $\bar{x}$ – среднее значение ряда x ;

$\sigma^2(x)$ – дисперсия ряда x ;

K – временной лаг.

Медленное убывание значений АКФ характеризует медленно убывающую зависимость между элементами трафика. Быстрое убывание значений АКФ – признак стационарности исследуемого процесса. Для фрактальных процессов, характеризующихся свойствами самоподобия и медленно убывающей зависимости, автокорреляционная функция не обращается в ноль, при $t \rightarrow \infty$.

Для более точного описания меры долговременной зависимости самоподобного процесса используют неоднородные фрактальные объекты, или мультифракталы [15]. При этом, исходный временной ряд $x(t)$ разбивается на N сегментов длиной S и для каждого сегмента

$$y(t) = \sum_{i=1}^n x(t),$$

Определяется функция

$$F(s) = \sqrt{\frac{1}{S} \sum_{t=1}^T (y(t) - y_m(t))^2}.$$

Далее находят зависимость $F_q(S)$ от фиксированного значения q

$$F_q(S) = \left\{ \frac{1}{N} \sum_{i=1}^N [F^2(S)]^{\frac{q}{2}} \right\}^{\frac{1}{q}}$$

Изменяя длину S временного ряда, при произвольных значениях q , находим последовательность значений $F_q(S)$. Если $F_q(S)$ представляется зависимостью [16]

$$F_q(S) \propto S^{h(q)},$$

то временной ряд соответствует мультифрактальному множеству и обладает долговременной зависимостью с показателем Херста $h(q) = H$.

Гармонический анализ.

Известно, что R/S -анализ не всегда дает корректные оценки показателей Херста. Большую статистическую точность дает метод, основанный на корреляционном анализе. Если справедливо выражение

$$\tau_k = k^{-\beta} L(k) + C,$$

где $C = \text{const}$, $0 < \beta < 1$;

$$L(k) - \text{медленно меняющаяся функция } \lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1.$$

то такой процесс описывается АКФ убывающую по степенному закону [12].

При описании сетевого процесса предпочтительным является использование гармонической функции вида

$$Y_t = \bar{Y}_t + \sum_{k=1}^M a_k \cdot \cos \frac{2\pi kt}{N} + \sum_{k=1}^M b_k \cdot \sin \frac{2\pi kt}{N} \quad (5)$$

где $a_k = \frac{2}{N} \sum_{k=1}^M Y_t \cdot \cos \frac{2\pi kt}{N}$,

$b_k = \frac{2}{N} \sum_{k=1}^M Y_t \cdot \sin \frac{2\pi kt}{N}$, – гармоники ряда Фурье;

M – количество гармоник ряда;

N – длина ряда.

Для описания долговременной тенденции временного ряда используем функцию [12]

$$\bar{Y}_T = A_1 \sin\left(\frac{2\pi t}{P}\right) + A_2 \cos\left(\frac{2\pi t}{P}\right) + A_3. \quad (6)$$

Период P цикла тренда получаем путем последовательного разбиения наиболее вероятного его значения на равные временные интервалы P_i и с помощью методов регрессии определяем коэффициенты A_1, A_2, A_3 . Далее выбираем последовательности временных интервалов в которых

$$\frac{1}{T} \sum_{t=1}^T (Y_T - F_{TP}(t, P_i))^2 \rightarrow \min.$$

Выбор модели, описывающей долговременную тенденцию сетевого процесса, производится на основе оценки её точности. Очевидно, что модели, обеспечивающие меньшее расхождение между реальными и расчетными значениями, дают большую точность. Наиболее целесообразно использовать в качестве показателей точности следующие выражения [17]

а) разность дисперсий $V_{ar} = \frac{1}{T} \sum_{t=1}^T (Y_t - F_{TP}(t))^2$;

б) ошибка приближения $A = \frac{1}{T} \sum_{t=1}^T \left| \frac{Y_t - F_{TP}}{Y_t} \right|$;

в) показатель детерминации процесса

$$R^2 = \frac{\sum_{t=1}^T (F_{TP}(t) - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2},$$

где $\bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t$ – среднее значение элементов временного ряда.

Теоретическое число частот выбирается равным $N/2$ при длине ряда, равной N . Практически требуются не все $N/2$, а лишь некоторые гармонические составляющие, выражающие основную часть вариации ряда. Для устранения случайных колебаний и возможных ошибок в расчетных значениях трафика полученный нестационарный временной ряд должен дополнительно обрабатываться. Возможным решением при этом является исключение его тренда, путем сглаживания динамики, например, методом центрированной скользящей средней, или скользящими параболами второго порядка, т.е. использовать абсолютные или нормированные отклонения временного ряда от его долговременной тенденции.

Для выделения регулярных периодических циклов, которые вносят наибольший вклад в общую динамику сетевого процесса, удаляем тренд из структуры временного ряда. Прогноз периодов циклов строится путем выделения гармоник спектра временного ряда с наибольшими амплитудами, формировании соответствующей модели и ее экстраполяции.

Функция, по которой рассчитывается Фурье-преобразование исследуемого временного ряда, имеет вид [18]

$$f(x) = \frac{a_0}{2} + \sum_{m=1}^M \left[a_m \cos\left(m \frac{2\pi}{X_{\max}}\right) + b_m \sin\left(m \frac{2\pi}{X_{\max}}\right) \right],$$

$$\text{где } a_m = \frac{2}{N} + \sum_{n=0}^{N-1} y_n \cos\left(m \frac{2\pi n}{N}\right), 0 \leq m \leq M;$$

$$b_m = \frac{2}{N} + \sum_{n=0}^{N-1} y_n \sin\left(m \frac{2\pi n}{N}\right), 0 \leq m \leq M;$$

M – количество гармоник исследуемого ряда;

(x_n, y_n) – массив выборки и значений временного ряда;

$n = 0, \dots, N$.

Известно [14], что для выделения из всех M гармоник временного ряда самых значимых, используется разложения ряда Фурье

$$f(x) = \frac{a_0}{2} + \sum_{m=1}^M \left[R_m \cos \theta_m \cos\left(m \frac{2\pi}{N}\right) + R_m \sin \theta_m \sin\left(m \frac{2\pi}{N}\right) \right],$$

$$\text{где } R_m = \sqrt{a_m^2 + b_m^2},$$

$$\theta = -\arctg \frac{b_m}{a_m}, -\pi \leq \theta_m \leq \pi;$$

$$a_m = R_m \cos \theta_m,$$

$$b_m = -R_m \sin \theta_m.$$

Обозначив частоту колебаний как $\omega_m = m/N$, распределение Фурье запишем в виде

$$f(x) = \frac{a_0}{2} + \sum_{m=1}^M R_m \cos(2\pi\omega_m n + \theta_m),$$

$$\text{где } T_m = \frac{1}{\omega_m} \text{ – период цикла временного ряда.}$$

Для получения численных значений амплитуды и фазы наиболее вероятных циклов временного ряда воспользуемся выражениями [13]

$$A_n = \frac{1}{n} \sqrt{R_e^2(S_n) + I_m^2(S_n)}$$

$$\varphi_n = \arctg \frac{I_m(S_n)}{R_e(S_n)},$$

где S_n – множество частот возможных циклов.

Тогда, описать цикл можно выражением

$$f_n(t) = A_n \cos(S_n t + \varphi_n).$$

Если представить динамику всплесков сетевого трафика в виде суммы множества циклических составляющих, то периодичность временного ряда будет равна

$$V(t) = \sum_D f_n(t),$$

где D – множество возможных циклов ряда.

Последовательность этапов выделения циклов временного ряда данных:

а) выбор исходного временного ряда и определение с помощью метода наименьших квадратов его тренда;

б) вычитание из исходного временного ряда полученного линейного тренда;

в) из полученного ряда выделить гармоник, обеспечивающих наибольший коэффициент детерминации

$$R^2 = 1 - \frac{\sum_{t=0}^{N-1} (y_t - Y_t^{garm})^2}{\sum_{t=0}^{N-1} (y_t - \bar{Y})^2}.$$

Известно [17], что для наиболее точных моделей коэффициент детерминации должен быть не меньше 0,8. В этом случае коэффициент корреляции близок к 1;

г) добавление выделенных гармоник в модель временного ряда;

д) проверка наличия автокорреляции с использованием показателей статистики Дарбина-Уотсона

$$d = \sum_{t=2}^n (y_t - y_{t-1})^2 / \sum_{t=1}^n y_t^2.$$

Значение $1,5 < d < 2,5$ является подтверждением допустимого качества модели регрессии;

е) определение краткосрочных циклов.

Полученная зависимость Y на интервале $t \in [0, N-1]$, дает возможность определить значения Y на интервале $t \in [N, N+m]$.

Использование данного подхода позволяет совершать достоверный краткосрочный прогноз и своевременно информировать систему балансировки ЦОД о последующих значительных выбросах сетевого трафика.

Нелинейное прогнозирование.

Пакетный трафик протокола TCP может быть описан не только как простой периодический процесс, но и как процесс, обладающий более сложным поведением и описываемый методами и моделями детерминированного хаоса. [19]. При этом, основным методом определения хаотичности данного нелинейного процесса является спектр характеристик, состоящий из n показателей Ляпунова. Знаки показателей Ляпунова достаточно надежно характеризуют тип колебаний исследуемого временного ряда. Детерминированный хаотический процесс определяется положительными показателями, нулевой показатель определяет квазипериодический процесс, отрицательный показатель представляет собой неподвижную точку фазовой траектории, называемую аттрактором системы. Иначе, все показатели Ляпунова детерминированного процесса отрицательны или равны нулю, у хаотического процесса хотя бы один положительный. Используя хаотические свойства исследуемого фрактального процесса, можно построить прогнозную модель его развития. В основе прогнозных моделей могут лежать методы линейного, нелинейного, прогноза, а также глобальной полиномиальной аппроксимации [20]. Для определения спектра нелинейных параметров временного ряда существует достаточно большой набор специального программного обеспечения. Например, пакеты программ CDA, Dataplore, RQA, TISEAN. Набор средств анализа, реализованный на платформе TISEAN, обеспечивает определение достаточно полного спектра нелинейных характеристик детерминированного хаотического процесса. При этом, хаотичность динамики процесса определяется значением максимального показателя Ляпунова, характеризующего скорость разбегания его фазовых траекторий. Для определения показателя Ляпунова используем программу lyap пакета TISEAN. В результате получаем зависимость коэффициента $S(\varepsilon, m, \Delta n)$ разбегания траекторий от времени [17]

$$S(\varepsilon, m, \Delta n) = \frac{1}{N} \sum_{n_0}^N \ln \left(\frac{1}{|U(S_{n_0})|} \times \sum_{S_n \in U(S_{n_0})} |S_{n_0+\Delta n} - S_{n+\Delta n}| \right),$$

где ε – окрестность точки S_{n_0} ;

m – фрактальная размерность фазового пространства;

Δn – интервал времени;

$U(S_{n_0})$ – окрестность точки S_{n_0} радиусом ε .

Полученные значения $S(\varepsilon, m, \Delta n)$, для различных Δn , характеризуют искомые значения показателей Ляпунова.

Общая модель нелинейного прогнозирования имеет вид [19]

$$x(t+T) = \frac{1}{|U_m|} \sum_{x(t') \in U_m} x(t'+T),$$

где U_m – окрестность точки $x(t)$.

Данный алгоритм реализован программами lzo-run, lzo-test, false-nearest платформы TISEAN.

При использовании линейного прогнозирования используется алгоритм [20]

$$x(t+T) = a_n x(t) + b_n,$$

$$\sum_{x(t') \in U_m} (x(t'+T) - a_n x(t') - b_n)^2 \rightarrow \min.$$

Прогнозный метод глобальной аппроксимации реализуется выражением [20]

$$\sum_t (x(t'+T) - f_T(x(t')))^2 \rightarrow \min$$

с помощью программы FNN платформы TISEAN.

Автоматизация процесса реализации прогнозной модели может быть проведена с помощью методики изложенной в [19], с помощью программ определения величины лага auto-cor.exe, выбора размера окна Тейлера mutual.exe, вычисления корреляционной размерности d2.exe.

Структура системы балансировки нагрузки.

Вычислительные ресурсы ЦОД представляются совокупностью кластеров серверов и средствами балансировки нагрузки. Структура системы балансировки нагрузки ЦОД показана на рисунке 2. По требованию пользователей система балансировки производит формирование виртуальных машин (ВМ) с определенными системными показателями. Локальный менеджер производит анализ загрузки серверов и размещение на них поступивших запросов. Система мониторинга ЦОД передает полученную информацию глобальному менеджеру, который осуществляет формирование ВМ, управляет балансировкой нагрузки кластеров серверов ЦОД, настраивает пропускную способность системы под профиль трафика. При этом последовательно реализуются следующие задачи:

- прием запросов от абонентов на реализацию услуг;
- расчет сетевой статистики, реализация алгоритма прогнозирования;
- распределение запросов по серверам кластера;
- выбор сервера в кластере, способного реализовать запросы пользователей;
- направление запросов к выбранному серверу кластера;
- распределение реализуемых программных приложений по серверам кластера;
- формирование совокупности виртуальных машин, реализующих приложения;
- прием результатов решения задач.

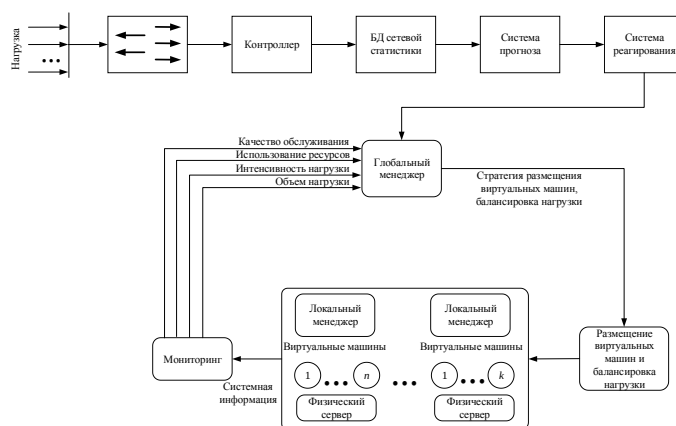


Рисунок 2. Структура системы балансировки нагрузки/ Fig. 2. The structure of the load balancing system

Для реализации алгоритма прогнозирования аномалий сетевой нагрузки ЦОД в информационную структуру системы балансировки дополнительно включается коммутатор, с подключенной по протоколу Gigabit Ethernet базой данных сетевой статистики с программой-сниффером WireShark на основе, например, библиотеки WinPcap, а также высокопроизводительный контроллер, реализующий алгоритм прогнозирования. Программа-сниффер осуществляет захват трафика, его обработку, агрегирование, формирование требуемых временных интервалов, а также фиксирует резкие изменения входной нагрузки. Полученный уровень всплесков интенсивности трафика, а также их длительность, информируют систему прогноза о необходимости включения на данном интервале агрегирования системы реагирования и перераспределения аппаратно-программных средств ЦОД. При этом очевидно, что запаздывание управляющего воздействия, связанное с процессом мониторинга сетевого трафика, может привести к снижению эффективности системы распределения и балансировки нагрузки. Поэтому необходимо осуществлять формирование превентивного управляющего воздействия, напрямую связанного с реализацией метода прогнозирования возможной пиковой нагрузки и времени ее появления.

Заключение

Экспериментальные исследования статистических характеристик пакетного трафика современных компьютерных сетей свидетельствуют о его фрактальной структуре, наличии частых всплесков и спадов активности, мощных пиковых выбросов, детерминированной составляющей. Подобные свойства трафика подтверждают возможность использования фрактальных моделей прогнозирования всплесков его интенсивности, определения с требуемой точностью объемов входящей нагрузки и решения на этой основе задач динамического управления системой распределения и балансировки нагрузки серверов кластеров ЦОД. Использование для расчета параметров системы балансировки нагрузки классических Марковских (без последствия) моделей и формул Эрланга, ориентированных на простейшие потоки, приводит, как правило, к некорректным, неэффективным для фрактальных (самоподобных) потоков результатам. В статье представлено описание метода балансировки нагрузки в основу которого положен достоверный, вероятностный подход упреждающего прогнозирования состояний сетевого трафика, сформированный на основе результатов его статистического и спектрального анализа, уровня фрактальности, плотности распределения. Задача кратковременного прогнозирования пакетного трафика сведена к задаче прогнозирования дискретного временного ряда. Разработанный метод способен обеспечить решение задачи эффективного планирования и распределения задач вычислительного кластера ЦОД с целью оптимизации использования ресурсов, ускорения времени выполнения задач и сокращения расходов на обработку приложений. Направлением дальнейших исследований, связанных с совершенствованием работы облачных ЦОД, является представление их в виде динамических систем, применение методов нелинейной динамики, позволяющих выявлять аттракто-

ры фазового пространства и обеспечивающих проведение более глубоких исследований влияния самоподобного трафика на производительность.

ЛИТЕРАТУРА

1. Tsybakov B.S., Georganas N.D. Self-similar processes in communications networks // IEEE Trans. Inform. Theory, vol. 44. Sep.2018. P. 1713-1725.
2. Tsybakov B.S., Georganas N.D. Self-similar traffic: upper bounds to buffer-overflow probability in an ATM queue // Proceedings of CCB'97, the Canadian Conference on Broadband Research, Ottawa. 2015. P. 137-148.
3. Gneiting T., Schlather M. Stochastic models which separate fractal dimension and Hurst effect // NRCSE-TRS. Sep. 20, 2001. № 069.
4. Feng W., Tinnakornsrisuphap P. The Failure of TCP in High-Performance Computational Grids // SC2000: High-Performance Network and Computing Conference, Dallas, TX. November, 2000.
5. Veres A., Boda M. The Chaotic Nature of TCP Congestion Control // Proceedings of IEEE INFOCOM'2000, March 2000.
6. Veres A., Kenesi Zs., Molnar S., Vattay G. On the Propagation of Long-Range Dependence in the Internet // Proc. ACM SIGCOMM 2000. Stockholm, Sweden, Sep. 2000.
7. Kugiumtzis D., Boudourides M. Chaotic Analysis of Internet Ping Data: Just a Random Generator? // SOEIS meeting at Bielefeld, March 27-28, 1998.
8. Park K., Willinger W. Self-Similar Network Traffic and Performance Evaluation. John Wiley & Sons, 2000.
9. Policing and Shaping Overview, QC: Cisco IOS Release 12.0 Quality of Service Solutions Configuration Guide. <http://www.cisco.com>.
10. Östring S., Sirisena H. The Influence the Long-Range Dependence on Traffic Prediction // Proceedings of ICC'01. Helsinki, June 2001.
11. Beran J. Statistical Methods for Data with Long-Range Dependence // StatisticalScience, Volume 7, Issue 4. 2019. P. 404-416.
12. Foag J., Wild T. Traffic Prediction Algorithm for a Speculative Network Processor // 17th Intl. Symposium for High Performance Computing Systems and Applications HPCS 2003. Sherbrooke, May 2003.
13. Trajkovic L., Neidhardt A. Internet traffic prediction // Centre for Systems Science, Simon Fraser University, Vol. 12, Issue 1. Mar. 2000.
14. Koucheryavy Y., Harju J. A novel approach for self-similar traffic prediction. / Proceedings of the St. Petersburg Regional International Teletraffic Seminar, St. Petersburg, Russia, January 29. February 1. 2002. P. 172-179.
15. H. Kantz and T. Schreiber. Nonlinear Time Series Analysis, 2nd edition, Cambridge University Press, Cambridge, 2003. – 388 p.
16. Fowler H.J., Leland W.E. Local area network traffic characteristic, with implications for broadband network congestion management // IEEE Journal on Selected Areas in Communications. vol. 9. 2021. P. 1139-1149.
17. Zhao H., Ansari N., Shi Y.Q. A Fast Non-linear Algorithm for Video Traffic Prediction // ITCC, 2002.
18. Chen B., Peng S., Wang K. Traffic Modeling, Prediction, and Congestion Control for High-Speed Networks: A Fuzzy AR Approach // IEEE Trans. On Fuzzy Systems Vol. 8. 2000. № 5.
19. Moon F.C. Chaotic and Fractal Dynamics: Introduction for Applied Scientists and Engineers. New York: Wiley; 1992.
20. Moon F. Chaotic Oscillations. Translate from English. Moscow: Mir Publ.; 1990. 312 p. (in Russian)

ОБ АВТОРАХ / ABOUT THE AUTHORS

Мочалов Валерий Петрович, профессор, доктор технических наук, профессор, профессор кафедры инфокоммуникаций, ФГАОУ ВО «Северо-Кавказский федеральный университет», институт цифрового развития, кафедра инфокоммуникаций, n.b.20062@yandex.ru, +7 9624004447, (8652) 95-69-97

Mochalov Valeriy Petrovich, Doctor of Technical Sciences, Professor, SI NCFU, Stavropol, Russia, n.b.20062@yandex.ru, +7 9624004447, (8652) 95-69-97

Братченко Наталья Юрьевна, кандидат физико-математических наук, доцент кафедры инфокоммуникаций, доцент, ФГАОУ ВО «Северо-Кавказский федеральный университет», институт цифрового развития, кафедра инфокоммуникаций, n.b.20062@yandex.ru, +7 9887405050, (8652) 95-69-97

Bratchenko Natalia Yurievna, candidate of physical and mathematical sciences, associate professor, SI NCFU, Stavropol, Russia, n.b.20062@yandex.ru

Палканов Илья Сергеевич, программист, ФГАОУ ВО «Северо-Кавказский федеральный университет», институт цифрового развития, кафедра инфокоммуникаций, n.b.20062@yandex.ru, +7 9289633235

Palkanov Ilya Sergeevich, programmer, SI NCFU, Stavropol, Russia, n.b.20062@yandex.ru, +7 9289633235

Дата поступления в редакцию: 12.03.2022

После рецензирования: 23.04.2022

Дата принятия к публикации: 13.06.2022