

## МЕТОД КОНТРОЛЯ КАЧЕСТВЕННОГО ИЗМЕНЕНИЯ СОСТОЯНИЙ СЕТЕВОГО ТРАФИКА НА ОСНОВЕ ИНФОРМАЦИОННОЙ МЕРЫ КУЛЬБАКА-ЛЕЙБЛЕРА

А.В. Скатков, А.А. Брюховецкий, Д.В. Моисеев

ФГАОУ ВО «Севастопольский государственный университет»

РФ, г. Севастополь, ул. Университетская, 33

E-mail: AVSkatkov@sevsu.ru, bryukhovetskiy@sevsu.ru, DVMoiseev@sevsu.ru

В работе рассматривается метод обнаружения изменений состояния сетевого трафика, в распределенных системах мониторинга состояния окружающей среды, базирующийся на основе информационной меры Кульбака-Лейблера.

**Ключевые слова:** мониторинг, сетевой трафик, J-эффект, система массового обслуживания, неоднородность наблюдений.

**Введение.** В настоящее время интенсивный рост требований к производительности, экономичности, доступности, и масштабируемости вычислительных систем, применяемых в системах мониторинга окружающей среды, приводит к необходимости развертывания высокопроизводительных распределенных вычислительных систем (РВС), в том числе облачных технологий обработки данных [1].

РВС применяются как в процессе оценки и прогнозирования измерений состояния окружающей среды, так и на этапе наблюдения [2].

Облачные вычисления представляют собой информационную технологию для обеспечения удобного сетевого доступа к комплексу измерительных приборов, объединенных информационными связями и предназначенными для совместного использования, по требованию, которые можно быстро активировать и предоставить с минимальными потерями на управление или минимальным взаимодействием с поставщиком услуг [1].

За последнее время в области мониторинга трафика был получен ряд важных теоретических и практических результатов. Так, например, в [3] предложена программная система *Unihub*, которая является облачной вычислительной системой типа *SaaS* и предоставляет пользователям возможность удаленной работы через *Web*-браузер с интерактивными графическими *Linux*-приложениями, которые поддерживают динамические требования к распределе-

нию вычислительных ресурсов. Такие решения позволяют выполнить балансировку имеющегося в распоряжении программно-аппаратного оборудования и увеличить предельную производительность системы мониторинга окружающей среды в целом.

В статье [4] рассматривается задача оценки объема трудозатрат на проектирование функциональных требований к информационной системе в сервис-ориентированной архитектуре. Для этого предлагается стохастическая модель процесса идентификации сервисов информационной системы, позволяющая при минимальных исходных данных дать оценку ожидаемого количества объектов проекта и связей между ними.

В своей работе [5] А.Н. Ширяев представляет вероятностно-статистические методы теории принятия решений, объясняющие, «как по наблюдениям за реализацией случайного процесса обнаружить спонтанно возникающие эффекты, когда свойства процесса изменяются скачкообразно». В условиях, в том числе, облачных вычислительных сред (ОВС) большая часть проблем анализа данных связана с исследованиями стохастических динамических систем, в которых обнаружение существенных, но редких информационных ситуаций часто имеет решающее значение.

Как правило, целями мониторинга современных систем могут быть: обеспечение масштабируемости и открытости, позволяющей управлять оборудованием различных производителей; де-

тальный анализ и исследование системы на предмет поиска узких, высоконагруженных мест; сбор информации об элементах сети, отслеживание системных событий и операций и др.

С целью развития известных подходов к решению указанных задач предлагается новый научный подход, который в значительной степени позволяет преодолеть требования к реализации известных методов контроля состояния сетевого трафика, в частности:

- больших объемов априорной информации для настройки параметров системы контроля,
- принятия решения о состоянии трафика в результате обработки большого объема текущих данных,
- трудно реализуемую стратегию обработки данных в реальном времени.

В силу отмеченных обстоятельств известные методы для решения задач мониторинга в ОВС не могут являться совершенным операционным средством, так как не обеспечивают требуемого качества контроля состояния сетевого трафика, в частности, с их помощью затруднительно отследить моменты изменения таких свойств трафика как интенсивность, дисперсность, нестационарность. С этой целью мы используем понятие дивергенции Кульбака-Лейблера (*J*-эффекта), которое используется при оценке однородности двух случайных выборок [6].

Основными преимуществами предлагаемого метода оценки состояний сетевого трафика на основе *J*-эффекта яв-

ляются:

- чувствительность к изменениям состояния трафика;
- малая вычислительная трудоемкость;
- адаптивность к внешним воздействиям.

Именно оценка неоднородности (назовем *J*-эффект, а устройство, которое будет распознавать *J*-эффект – *J*-детектор) состояния сетевого трафика может использоваться для обнаружения внешних воздействий, каковыми являются аномальные явления (А-события), или изменения состояния сетевого трафика, которые могут быть вызваны изменениями в спросе-предложении участников рынка сервисных услуг.

Результаты экспериментов по распознаванию *J*-эффекта в узле сети на примере СМО. Процесс обнаружения *J*-эффекта в узле сети многоэтапный и включает в себя: задание закона распределения  $\lambda$  входного потока; задание интервалов изменения  $\lambda$ ; вычисление выходных характеристик: процент потерянных заявок, время обработки заявок как функции от  $\lambda$  и  $\mu$ ; построение регрессионных зависимостей; сбор статистических данных о динамике изменения дисперсий остатков; принятие решения о наличии *J*-эффекта на основе дивергенции Кульбака-Лейблера.

Для мониторинга качественного изменения состояний сетевого трафика в расширении *Simulink* пакета *MATLAB* была построена модель СМО (рис. 1).

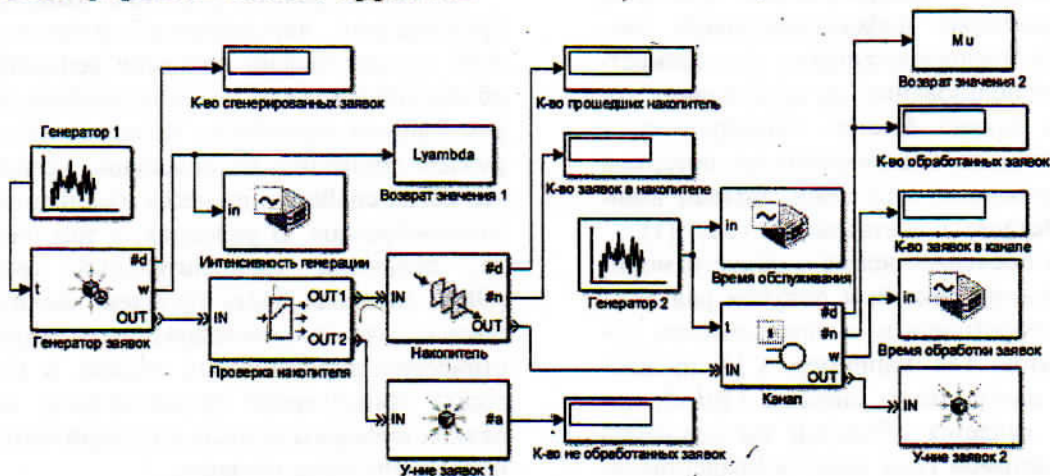


Рис. 1. Модель СМО построенная в расширении *Simulink* пакета *MATLAB*

Моделируемая СМО содержит следующие основные блоки:

- генератор заявок, подключенный к генератору нормального закона распределения, выполняющего функцию генерации заявок с заданными  $M(\lambda)=0,7$ ,  $\sigma(\lambda)=0,086$ ;
- накопителя заявок ёмкостью 5, с алгоритмом функционирования FIFO;
- блока проверки занятости накопителя;
- канала обслуживания заявок, с подключенным генератором нормального закона распределения, выполняющего функцию генерации времени обработки заявок с заданными параметрами  $M(\mu)=0,7$  и  $\sigma(\mu)=[0,17; 0,2; \dots; 0,43]$  в соответствии с экспериментом;
- блоки индикации, предназначенные для вывода результатов моделирования:
  - индикатор количества сгенерированных заявок;
  - индикатор количества заявок побывавших в накопителе;
  - индикатор количества заявок находящихся в настоящий момент в накопителе;
  - индикатор количества обработанных заявок;
  - индикатор количества заявок попавших в отказ обработки;
  - индикатор количества заявок находящихся в канале в настоящий момент времени;
  - блоки возврата промоделированных значений  $\lambda$  и  $\mu$  из *Simulink* в *MATLAB*;
  - блоки построения графиков для  $\lambda$  и  $\mu$ ;
  - блоки уничтожения заявок (обработанных и попавших в отказ).

Расчет дивергенции Кульбака-Лейблера выполнялся в соответствии с формулой [5]:

$$J = \sum_{i=1}^k \left( \frac{f_i}{m_1} - \frac{q_i}{m_2} \right) \ln \frac{f_i m_2}{q_i m_1}, \quad (1)$$

где  $f_i$ ,  $q_i$  – числа попадания в  $i$ -й интервал примеров сравниваемых выборок,  $m_1$ ,  $m_2$  – численность обеих выборок.

Цель эксперимента: определить условия возникновения  $J$ -эффекта, т.е. при заданных  $M(\lambda)$ ,  $\sigma(\lambda)$ ,  $M(\mu)$ , варьируя  $\sigma(\mu)$ , определить, в частности, устойчивый порог срабатывания детектора  $J$ -эффекта для уровня значимости  $p$ . Функции распределения – нормальные.

Результаты моделирования построенной СМО представлены в табл. 1 и содержат десять экспериментов выполненных для заданных значений  $M(\lambda)=0,7$ ;  $\sigma(\lambda)=0,086$ ;  $M(\mu)=0,8$  и варьируемым значением  $\sigma(\mu)$  в диапазоне  $\sigma(\mu)=[0,17\dots 0,43]$  с шагом 0,03.

Проведены эксперименты и получены результаты в среде имитационного моделирования *MATLAB*, которые подтверждают факты наличия возмущений в моделируемом трафике, в который привносится измененный входной поток заявок  $\xi$  и/или измененная интенсивность обработки заявок.

В результате каждого эксперимента формировался массив из 100 значений  $\lambda$  и  $\mu$ . В качестве примера в табл. 1 представлены значения первого эксперимента с параметрами  $M(\lambda)=0,7$ ;  $\sigma(\lambda)=0,086$ ;  $M(\mu)=0,8$ ;  $\sigma(\mu)=0,17$ .

Таким образом, полученные в экспериментах значения  $\mu$  разбивались на 15 интервалов с шагом  $\pm 0,05$ .

На рис. 2 приведено количество попаданий  $\mu$  в  $i$  ( $i=1, 2 \dots 15$ ) диапазон в  $j$ -ом эксперименте ( $j=1, 2 \dots 5$ ).

Рассчитав количество попаданий значения  $\mu$  в каждый из 15 поддиапазонов для всех 10 экспериментов можно воспользоваться формулой (1) для расчета дивергенции Кульбака-Лейблера. В нашем случае количество выборок  $m_1=m_2=100$ ,  $k=15$ .

Таблица 1. Значения результатов первого эксперимента с параметрами  $M(\lambda)=0,7$ ;  $\sigma(\lambda)=0,086$ ;  $M(\mu)=0,8$ ;  $\sigma(\mu)=0,17$

№	$\lambda$	$\mu$	№	$\lambda$	$\mu$	№	$\lambda$	$\mu$
1	0,4381	0,4428	34	0,1286	0,5444	67	0,2641	0,3241
2	0,4584	0,4323	35	0,0855	0,4981	68	0,0542	0,4364
3	0,4319	0,2426	36	0,0855	0,5722	69	0,3418	0,2835
4	0,4259	0,3413	37	0,1675	0,2223	70	0,2547	0,9890
5	0,4259	0,9952	38	0,2628	0,4251	71	0,2193	0,3288
6	0,3931	0,6444	39	0,1220	0,2788	72	0,3171	0,4897
7	0,2663	0,2946	40	0,3164	0,3391	73	0,4410	0,6616
8	0,2755	0,1462	41	0,3106	0,1869	74	0,4410	0,1716
9	0,3016	0,0174	42	0,1328	0,1869	75	0,0742	0,9758
10	0,1547	0,0609	43	0,2367	0,6105	76	0,0742	0,4106
11	0,0694	0,7555	44	0,2367	0,3849	77	0,3011	0,4942
12	0,0694	0,1023	45	0,4073	0,5684	78	0,4346	0,6756
13	0,0694	0,0159	46	0,1418	0,0901	79	0,1231	0,0155
14	0,0031	0,2993	47	0,4984	0,9602	80	0,2382	0,5801
15	0,1756	0,7670	48	0,1222	0,3792	81	0,1437	0,9545
16	0,1756	0,2197	49	0,2344	0,7057	82	0,1437	0,1615
17	0,0907	0,1212	50	0,2690	0,4009	83	0,0035	0,7514
18	0,2556	0,2511	51	0,0377	0,0523	84	0,1719	0,4326
19	0,4032	0,5126	52	0,3584	0,7335	85	0,1480	0,4237
20	0,4032	0,0908	53	0,3584	0,2150	86	0,3704	0,0408
21	0,2474	0,0333	54	0,3390	0,3815	87	0,1718	0,8943
22	0,2655	0,9370	55	0,1849	0,9891	88	0,0646	0,1774
23	0,2065	0,7118	56	0,1041	0,9459	89	0,0174	0,0095
24	0,0823	0,6469	57	0,1041	0,3385	90	0,2666	0,3401
25	0,2301	0,2938	58	0,4906	0,3457	91	0,0264	0,9409
26	0,2655	0,5403	59	0,0264	0,8726	92	0,0264	0,5230
27	0,1641	0,4414	60	0,4871	0,3436	93	0,4546	0,1620
28	0,4889	0,3822	61	0,1481	0,7941	94	0,1996	0,1181
29	0,2147	0,1133	62	0,1481	0,4394	95	0,1776	0,7223
30	0,3800	0,3963	63	0,1073	0,3652	96	0,4218	0,9434
31	0,2466	0,0076	64	0,1164	0,2915	97	0,1166	0,5889
32	0,2167	0,4894	65	0,3181	0,5347	98	0,4322	0,7321
33	0,3541	0,2172	66	0,3181	0,2932	99	0,0264	0,7769
						100	0,4736	0,2170

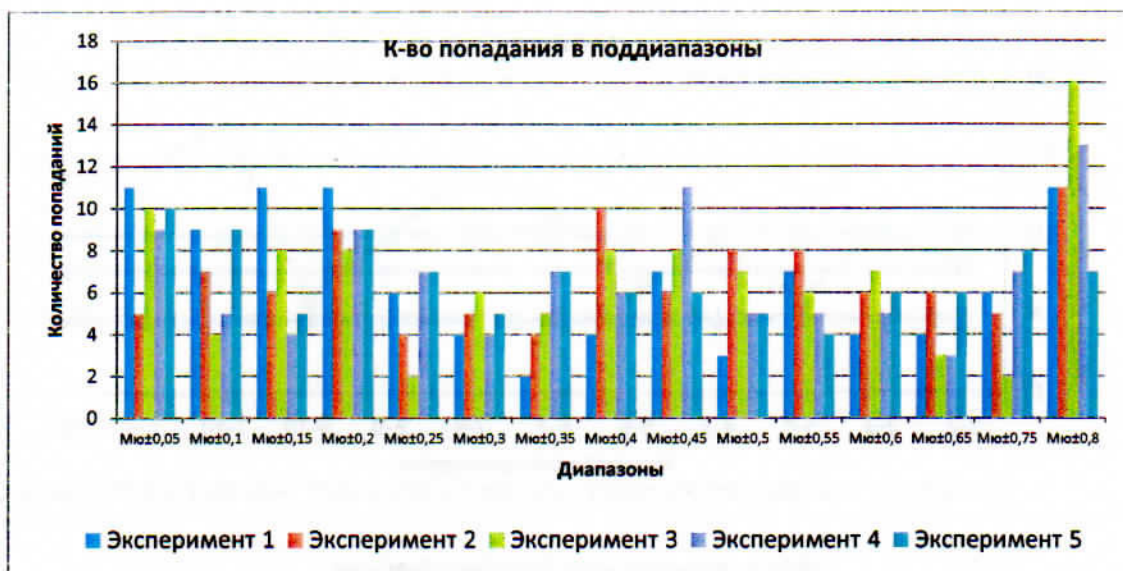


Рис. 2. Количество попаданий  $\mu$  в  $i$ -й диапазон  $j$ -го эксперимента

В табл. 2 представлены рассчитанные значения меры и дивергенции Кульбака-Лейблера, значения  $\chi^2$  для доверительных вероятностей  $p=0,99$ ,  $p=0,9$ ,  $p=0,75$ , полученные при  $r=14$  степеней свободы.

Для расчета мера Кульбака-Лейблера использовалась следующая формула [5]:

$$J' = \frac{m_1 \cdot m_2}{m_1 + m_2} \cdot J.$$

В соответствии с данными, приведенными, в табл. 2 был построен график зависимости меры Кульбака-Лейблера от номеров экспериментов (рис. 3).

Как видно из рис. 3 при значении доверительной вероятности  $p=0,75$   $J$ -эффект становится особенно сильно заметным при сравнении выборок первого с десятым, второго с девятым и третьего с восьмым экспериментов, что является вполне закономерным т.к. значения  $\sigma(\mu)$  в этих случаях максимально отличаются друг от друга.

При увеличении значения доверительной вероятности до  $p=0,99$   $J$ -эффект становится заметным даже для выборок, в которых значение  $\sigma(\mu)$  отличается лишь на 0,03.

Таблица 2. Значения меры и дивергенции Кульбака-Лейблера

Значения	Номера сравниваемых экспериментов						
	1-2	2-3	3-4	4-5	5-6	6-7	7-8
$J$	0,162	0,209	0,123	0,190	0,133	0,231	0,294
$J'(14,1)$	8,108	10,456	6,133	9,494	6,648	11,541	14,690
$\chi^2$ $p=0,99$ , $r=15-1$	4,660	4,660	4,660	4,660	4,660	4,660	4,660
$\chi^2$ $p=0,9$ , $r=15-1$	7,790	7,790	7,790	7,790	7,790	7,790	7,790
$\chi^2$ $p=0,75$ , $r=15-1$	10,165	10,165	10,165	10,165	10,165	10,165	10,165
Значения	8-9	9-10	1-10	2-9	3-8	4-7	
$J$	0,200	0,121	0,349	0,287	0,281	0,241	
$J'(14,1)$	10,010	6,055	17,437	14,349	14,054	12,073	
$\chi^2$ $p=0,99$ , $r=15-1$	4,660	4,660	4,660	4,660	4,660	4,660	
$\chi^2$ $p=0,9$ , $r=15-1$	7,790	7,790	7,790	7,790	7,790	7,790	
$\chi^2$ $p=0,75$ , $r=15-1$	10,165	10,165	10,165	10,165	10,165	10,165	

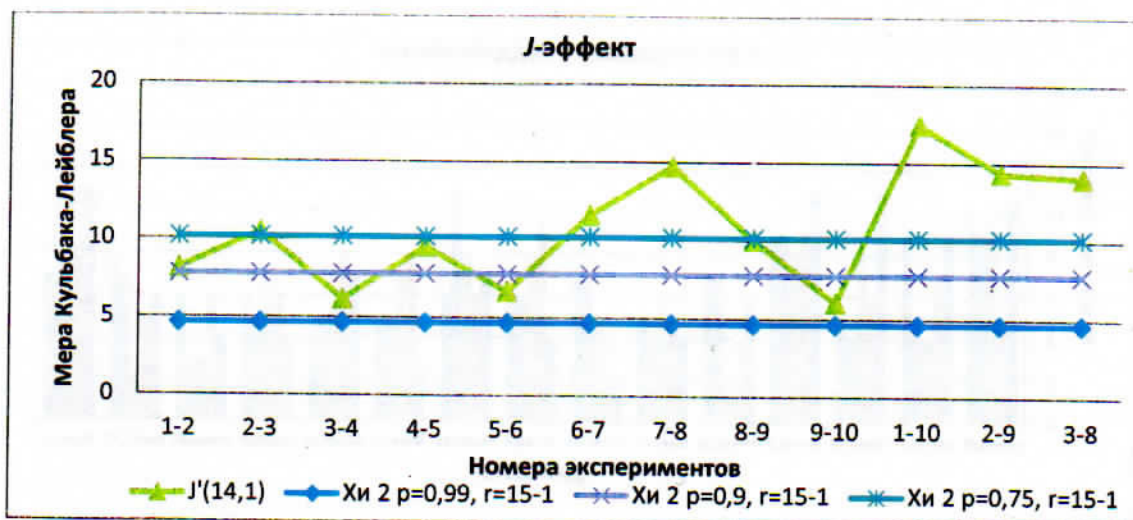


Рис. 3. Значения меры Кульбака-Лейблера

**Заключение.** Проведены эксперименты и получены результаты в среде имитационного моделирования *MATLAB*, которые подтверждают факты наличия возмущений в моделируемом трафике, системы мониторинга состояния окружающей среды, в который приносится измененный поток интенсивности обработки заявок.

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований (грант № 15-29-07936).

#### СПИСОК ЛИТЕРАТУРЫ

1. Mell P., Grance T. The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology, 2010.

2. Якунина И.В. Методы и приборы контроля окружающей среды. Экологический мониторинг: уч. пос. Тамбов: Изд-во Тамб. гос. техн. ун-т, 2009. 188 с.

3. Грушин Д.А., Кузюрин Н.Н. Балансировка нагрузки в системе Unihub на основе предсказания поведения пользователей // Труды Института системного программирования РАН, 2015. Том 27. Вып. 5. С. 23–34.

4. Циперман Г.Н. Стохастическая модель процесса идентификации сервисов информационной системы // Труды ИСП РАН, 2014. Том 26. Вып. 5. С. 7–28.

5. Ширяев А.Н. Вероятностно-статистические методы в теории принятия решений. 2-е изд., новое. М.: МЦНМО, 2014. 144 с.

6. Генкин Б.М. Экономика и социология труда: учебник для вузов, 1998. 384 с.

#### QUALITY CONTROL METHOD ALTERED STATE OF NETWORK TRAFFIC BASED ON INFORMATION CONTENT OF KULLBACK-LEIBLER

A.V. Skatkov, A.A. Brjuhoveckij, D.V. Moiseev

Federal State Educational Institution of Higher Education «Sevastopol State University»  
Russian Federation, Sevastopol, Universitetskaya St., 33

In this paper the method of detecting changes in the state of network traffic in distributed systems for monitoring the state of the environment, based on the basis of the information measures the Kullback-Leibler.

**Keywords:** Monitoring, network traffic, *J*-effect, queuing system, the heterogeneity of observations.