УДК 004.75

СТАТИСТИЧЕСКИЕ ОЦЕНКИ РИСКОВ В УСЛОВИЯХ НЕСАНКЦИОНИРОВАННЫХ ВОЗМУЩЕНИЙ УЗЛОВОГО ТРАФИКА

А.В. Скатков, К.С. Ткаченко

ФГАОУ ВО «Севастопольский государственный университет»
РФ, г. Севастополь, ул. Университетская, 33
E-mail: itiks@sevsu.ru

Предлагается подход по оценке характера функционирования узлов распределенных сред. В основе подхода лежит представление узлов в виде моделей и использование непараметрических статистик выборок. Применение подхода позволяет эффективно и результативно оценивать гипотезы о параметрах трафика в узле. Это позволяет построить управление отдельным узлом и всей распределенной средой в целом.

Ключевые слова: трафик узла, распределенная среда, статистические методы.

Введение. Современное проактивное управление критическими объектами основано не только на планировании и регулировании, но и в большей мере на мониторинге объектов критического применения [1–4]. В таких объектах заранее известны составляющие их элементы, коммуникации взаимодействия между ними, а также характерны устойчивые структуры. Например, в современном мире технические решения основаны на доминантном применении распределенных сред (РС) – системы управляющих контуров средств вычислительной техники систем контроля окружающей среды, крупнейшие системы хранения данных, и прочего.

Синтез, анализ и исследование структуры информационного, программного и математического обеспечения достаточно трудоемко само по себе и в настоящей работе не рассматривается. Предметом рассмотрения является статистическое оценивание важнейших системных узловых характеристик. Предположим, что известны основные характеристики топологической и технической структур. Зная их, можно вычленить отдельные элементы РС, а именно, узлы. Характеристики таких узлов изменяются в зависимости от входных факторов, в качестве которых используются среднее число заявок в очереди, средняя загрузка системы и некоторые другие.

Рассмотрим достаточно частный случай модели узла – систему массового обслуживания (СМО) M/M/K/N – то для него таковыми станут число каналов K, емкость буфера N, интенсивность входного потока заявок λ, производительность устройств обработки μ.

Узел предназначен для обработки в нормальных условиях некоторого полезного трафика с интенсивностью λ. Но вследствие использования средств мониторинга, вирусных атак, несанкционированного доступа происходят его флуктуации, чаще всего в большую сторону $\lambda^{(*)}$. Тогда можно говорить о так называемой несанкционированных возмущениях трафика, в дальнейшем именуемыми В-событиями.

В-события приводят к различным рискам. По одному из возможных определений риск численно оценивают как произведение стоимость компенсации последствий негативного события на вероятность его появления. Эта стоимость определяется пользователем, а сама задача является внешней. Ниже в настоящей работе будет рассматриваться задача статистической оценки вероятности. Эту сложную, нерешенную в общем виде задачу предлагается решать методами статистического моделирования.

Предельным случаем такого риска является риск потери заявок. Он проявляется при предельном заполнении входного буфера. Стоит отметить, что в зависимости от кортежа $<\lambda, \mu>$ изменяются отклики СМО при фиксированном кортеже $<K, N>$. Тогда и только тогда становится возможным оценить интервальный дрейф отклика «среднее число заявок в СМО». Это можно выполнить

известными методами непараметрической статистики.

Целью проводимых далее исследований является построение способов оценки узлового риска непараметрическими статистиками.

Информационный обзор современного состояния проблем. В фундаментальном труде [1] заложены основы теоретического проактивного мониторинга и управления сложными объектами, широко описаны технологические основы проактивного мониторинга и управления. Эти положения необходимы для описания функционирования сложных систем на модельном и алгоритмическом уровне, а затем, на основе описания, решить задачи комплексной автоматизации. Недостатками частичного решения указанных задач являются ручное выполнение управляющих воздействий и оценка обстановки. Для выполнения интегральной оценки предлагается использование оперативного формирования процедур сбора, обработки, анализа данных.

Известны системы адаптивного распределения ресурсов в интегрированных сетях ресурсного снабжения [2]. Такие системы в идеальном случае должны обеспечить полноценный мониторинг инфраструктур сложных объектов, для последующего эффективного распределения ресурсов. Быстросходящаяся адаптация возникает на возможном прогнозировании и работами в соответствии с планом. При этом используются принципы конкуренции и кооперации на основе мультиагентных технологий.

Динамический синтез ограничений в процессе имитации и разработка математического аппарата для унификации обработки данных и знаний позволяет строить концептуальные модели [3]. Для этих целей осуществляется полный перебор функциональных отображение модели. Условием завершения итеративного процесса является не только получение всех необходимых результатов, но и при отсутствии требуемого отображения.

Состав и структура запросов к базе моделей и полимодельных комплексов определяется на основе обобщенной структуры выбора альтернатив [4]. Применяя концепции структурно-математического и категорийно-функторного подходов формулируются и решаются задачи выбора управляющих воздействий в автоматизированных системах управления.

Широкоадресная передача данных иногда может существенно уменьшить трафик в сетях передачи данных, но ее применение ограничено вследствие проблем управления, возникающих при масштабировании [5]. Поэтому предлагаются методы коммутации на основе одноадресного переключения стратегий. Решение позволяет организовывать программно-определяемые сети. Для снятия ограничений на задержку предлагается механизм активной вставки правил. Все это позволяет уменьшить требования к ресурсам и проектировать следующие поколения компьютерных сетей.

При использовании человека в контуре обратной связи управления необходимо выработать метрики качества управления с ним [6]. Настраиваемые коэффициенты определяются на основе требований пользователя при условии ограничений на ресурсы. Измеряемые характеристики системы определяются на основе запросов и их корреляции, и включают себя затраты на гибкие действия.

Рассмотрены пути ограничения роли операторов центров обработки данных путем прогностической модели для сбоев узлов [7]. Состояние узлов в течение длительного времени описывается ансамблем случайных классификаций.

Существуют описанные процессы для оценки рисков информационной безопасности, получения их численной оценки с целью принятия эффективных мер по защите информации без привлечения экспертов [8].

Проработана техническая концепция анализа риска [9], в основе которой лежит анализ относительных частот возникновения опасных явлений или их последствий как способе задания их вероятностей.

Таким образом, сделано многое для выработки решений с ЛПР по оценке качества результатов структурного синтеза. Слабо представлено использование методов непараметрической статистики

для факторной оптимизации по узловым системным откликам.

Методы непараметрической статистики из готовых программных пакетов. Широко известны и наиболее распространены два метода непараметрической статистики – простой критерий знаков (sign test) и критерий Уилкоксона (Willcoxon test, в некоторых источниках Уилкоксона). Сильно упрощая, эти методы заключены в следующем [10].

Если выполнять фиксирование откликов системы до и после управляющего воздействия, либо до и после воздействия типа несанкционированного изменения трафика, то получаются так называемые парные измерения $<a_i, b_i>$, где a_i – до воздействия, b_i – после, i – номер измерения. Разница $b_i - a_i$ является знаковым вещественным числом, так называемым «сдвигом». С помощью этих самых критериев можно проверить гипотезу H0 = {нет сдвига}.

Критерий знаков работает с переменной-счетчиком и не рассматривает нулевые разности. Счетчик увеличивается при положительной разности. Вероятность гипотезы о значимости сдвига и направления рассчитывается по соотношениям для испытаний Бернулли.

В отличие от него, знаково-ранговый критерий Уилкоксона учитывает ранг разностей. При упрощенной проверке критическое значение, получаемое по критерию, есть минимальная величина из сумм произведений рангов на счетчики.

Методы достаточно легко реализуются программно, в том числе на сценарных языках программирования и в табличных процессорах офисных пакетов. Стоит отметить, что в настоящей работе применяется третий способ, а именно, задействование готовых программных пакетов математической статистики, содержащих необходимый функционал непараметрической статистики. Более того, такие пакеты имеют встроенные трансляторы для языков программирования высокого уровня, которым отводится роль макроавтоматизации. Это избавляет оператора-пользователя от выполнения рутинных действий.

Определение вероятностей гипотез. Полагается, что H0 – трафик невозмущен, H1 – трафик возмущен. Тогда условные вероятности гипотез: P(H0|H0) – предполагается невозмущенный трафик, если он невозмущен; P(H0|H1) – предполагается возмущенный трафик, если он невозмущен; P(H1|H0) – предполагается невозмущенный трафик, если он возмущен; P(H1|H1) – предполагается возмущенный трафик, если он возмущен. Существуют законы распределения вероятностей, такие как равномерный, нормальный, экспоненциальный, гамма.

Расчет оценок проводится по двум критериям – знаков и Уилкоксона. Число выборок N = 20. Объем выборки k = 100. Вводится термин группа выборок: I группа – отсутствуют возмущения; II группа – возмущения линейно возрастают в зависимости от номера выборки. Будет проводится проверка I группы с I группой для расчета P(H0|H0) и P(H1|H0), затем – I группы со II Группой для расчета P(H0|H1) и P(H1|H1). Это вызвано тем, что критерии не сравнивались для одних и тех же выборок-«переменных». Сравнение производится для выборок I группы с I группой и со II группой, потому что I группа являлась эталонной. Не имеет смысла сравнивать II группу со II группой.

Пусть P(H0|H0)=a1/b1, P(H1|H0)=a2/b1, P(H0|H1)=a3/b2, P(H1|H1)=a4/b2,

где a1 – число предположений-невозмущений при невозмущении (определяется при I с I);

a2 – число предположений-возмущений при невозмущении (определяется при I с I);

a3 – число предположений-невозмущений при возмущении (определяется при I со II);

a4 – число предположений-возмущений при возмущении (определяется при I со II). Общее число сравнений для каждого критерия (как знаков, так и Уилкоксона):

- Для сравнения групп I-й с I-й есть b1 = (N/2–1) + (N/2–2) + … + (N/2–10) = 5*(N–11) = 45.
- Для сравнения групп I-й со II-й есть b2 = (N/2) * (N/2) = N^2/4 = 100.

Номер выборки: 1, 2, …, N.

В результате проведенных вычислительных экспериментов строится таблица принятия решений (табл. 1).

Таблица 1. Таблица принятия решений

№	Критерий	Закон распределения	Число выборок	Уровень дост.	P(H0|H0)	P(H1|H0)	P(H0|H1)	P(H1|H1)
1	Знаков	Равномерный	10 невозм. + 10 возм.	0,95	1,0000	0,0000	0,1600	0,8400
2	Уилкоксона	Равномерный	10 невозм. + 10 возм.	0,95	0,9778	0,0222	0,1400	0,8600
3	Знаков	Нормальный	10 невозм. + 10 возм.	0,95	0,9700	0,0300	0,9900	0,0100
4	Уилкоксона	Нормальный	10 невозм. + 10 возм.	0,95	0,9111	0,0889	0,9700	0,0300
5	Знаков	Экспоненциальный	10 невозм. + 10 возм.	0,95	0,8889	0,1111	0,0400	0,9600
6	Уилкоксона	Экспоненциальный	10 невозм. + 10 возм.	0,95	0,8667	0,1333	0,0000	1,0000
7	Знаков	Гамма	10 невозм. + 10 возм.	0,95	1,0000	0,0000	0,1400	0,8600
8	Уилкоксона	Гамма	10 невозм. + 10 возм.	0,95	1,0000	0,0000	0,1100	0,8900

Выбор критерия – знаков или Уилкоксона – можно произвести, основываясь на так называемых кривых насыщения. С этой целью необходимо построить графики зависимостей частоты обнаружения в зависимости от сдвига. При этом левая граница интервала для генератора псевдослучайной последовательности с непрерывным законом распределения равномерно возрастает. В каждом случае происходит сравнение 1-й выборки со 2-й, 3-й, …, 21-й с использованием критерия знаков и критерий Уилкоксона. Подсчитывается количество обнаружений и необнаружений возмущений. Результаты приводятся на рисунках 1 и 2.

Рис. 1. Зависимость частоты обнаружений возмущений от сдвига

Рис. 2. Зависимость частоты необнаружений возмущений от сдвига

Вероятности состояний СМО. Проводятся вычислительные эксперименты по определению линии разделения облаков вероятностей состояний СМО типа M/M/K/N. Для этого достаточно выделить два состояния, которые можно назвать p0 и p1. p0 – это вероятность простоя системы,

$$p0 = \left[1 + \sum_{j=1}^{K-1} \frac{\rho^j}{j!} + \frac{\rho^K (1-\rho_c^{N+1})}{K!(1-\rho_c)}\right]^{-1}.$$

p1 — это вероятность загрузки системы, p1 = 1,0 – p0. В этих формулах $\rho = \frac{\lambda}{\mu}$, $\rho_c = \frac{\lambda}{K\mu} = \frac{\rho}{K}$. Порождаются векторы-

столбцы $\langle \lambda^0, \mu^0 \rangle$, $\langle \lambda^1, \mu^1 \rangle$ по формулам для заданных K, N, i=0,1,…$i^{(max)}$, λ_i^0=w_1+w_2i, μ_i^0=w_3+w_4i, λ_i^1=w_5+w_6i, μ_i^1=w_7+w_8i.

В частности, для случая K=3, N=5, i=0,1,…,49: λ_i^0=25+0,5i, μ_i^0=10+0,5i, λ_i^1=30+0,5i, μ_i^1=20+0,5i получается графики разделенных облаков, изображенные на рис. 3. На рисунке зависимость для $\langle \lambda^0, \mu^0 \rangle$ изображается непрерывной линией, $\langle \lambda^1, \mu^1 \rangle$ – пунктирной. Облака вероятностей состояний СМО сужены в линии.

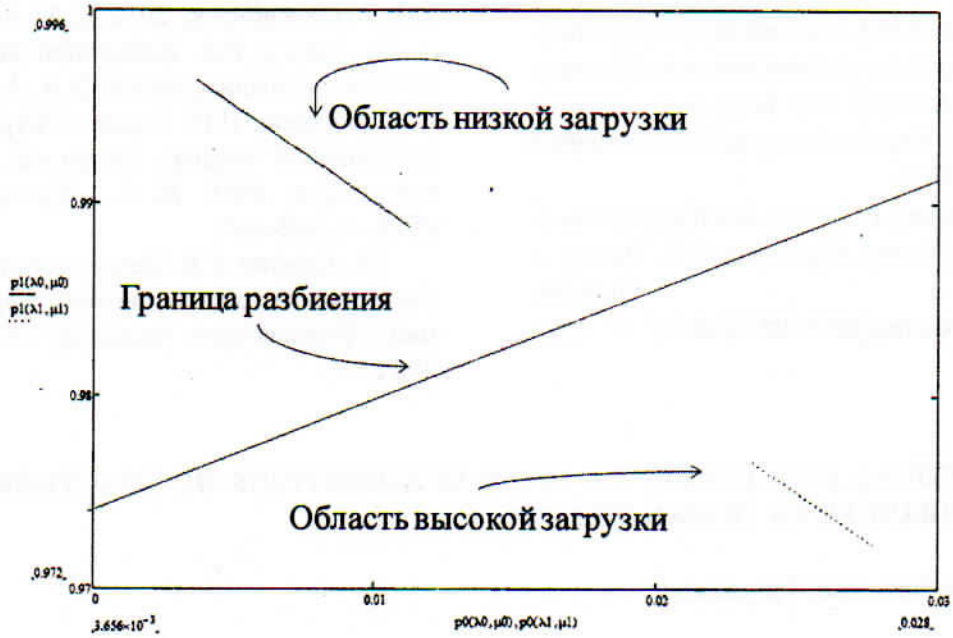

Рис. 3. Пример разделения облаков вероятностей состояний СМО

Заключение. В статье приводится принципиально новый подход к оценке характера функционирования узлов распределенных сред. Проблемная ситуация состояла в невозможности определения предельных значений условных вероятностей гипотез. Метод ее разрешения на основе статистического оценивания позволяет эффективно и результативно оценивать гипотезы о параметрах трафика в узле. Это дает возможность построения системы поддержки принятия решений по управлению как отдельным узлом, так и всей РС в целом.

Работа выполнена в рамках задания по базовой части госзаказа темы № 3866, а также при частичной поддержке Российского Фонда Фундаментальных исследований, грант №15-29-07936.

СПИСОК ЛИТЕРАТУРЫ

1. *Охтилев М.Ю.* Теоретические и технологические основы концепции проактивного мониторинга и управления сложными объектами / М.Ю. Охтилев, Б.В. Соколов, Р.М. Юсупов // Известия ЮФУ. Технические науки. С. 162–174.

2. *Будаев Д.С.* Разработка интеллектуальной сетецентрической системы адаптивного распределения ресурсов в интегрированных сетях газо-, тепло- и электроснабжения / Д.С. Будаев, В.Б. Ларюхин, Д.С. Косов, Е.В. Симонова // Вестн. Самар. гос. техн. ун-та. Сер. технические науки. 2013. № 3 (39). С. 6–14.

3. *Зуенко А.А.* Управление ограничениями в системах концептуального мо-

делирования: имеющийся задел и перспективы / А.А. Зуенко, А.Я. Фридман // Труды Кольского научного центра РАН, 4/2011 (7). Апатиты. 2011. С. 120–127.

4. *Губанов Н.Г.* Концепция разработки информационной системы поддержки принятия решений при управлении сложными техническими системами / Н.Г. Губанов, А.В. Чуваков // Вестн. Самар. гос. техн. ун-та. Сер. технические науки. 2013. № 3 (39). С. 21–31.

5. *Reed M.J.* Stateless multicast switching in software defined networks [Электр. текст. данные] / M.J. Reed et al. – Режим доступа: http://arxiv.org/pdf/1511.06069v1 19 Nov 2015.

6. *Huang L.* System Intelligence: Model, Bounds and Algorithms / L. Huang. – Режим доступа: http://arxiv.org/pdf/1605.02585v1 9 May 2016.

7. *Sirbu A.* Towards Operator-less Data Centers Through Data-Driven, Predictive, Proactive Autonomics / A. Sirbu, O. Babaoglu. – Режим доступа: http://arxiv.org/pdf/1606.04456v1 14 Jun 2016.

8. *Плетнев П.В.* Методика оценки рисков информационной безопасности на предприятиях малого и среднего бизнеса / П.В. Плетнев, В.М. Белов // Доклады ТУСУР, № 1 (25), часть 2, июнь 2012. Новосибирск. 2012. С. 83–86.

9. *Акимов В.А.* Концепции риска и концепции анализа риска / В.А. Акимов, С.П. Воронов, Н.Н. Радаев // Стратегия гражданской защиты: проблемы и исследования, 2013. № 2. Красноярск. 2013. С. 562–567.

10. *Хиценко В.Е.* Непараметрическая статистика в задачах защиты информации // Новосибирск: Изд-во НГТУ. 2012. 196 с.

STATISTICAL EVALUATION OF RISKS IN CONDITIONS OF UNAUTHORIZED DISTURBANCES OF NODAL TRAFFIC

A.V. Skatkov, K.S. Tkachenko

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

We propose an approach for the assessment of the functioning of the nodes in distributed environments. The approach is based on the representation of the nodes in the form of models and the use of non-parametric statistics of samples. The approach allows efficient and effective to assess the hypotheses about the node traffic parameters. This allows you to build a separate management node and all distributed environments in general.

Keywords: traffic node, distributed environment, statistical methods.