УДК 004.9
DOI 10.17513/snt.39946

КЛАССИФИКАЦИЯ ДАННЫХ РАСПРЕДЕЛЕНИЯ ВРЕМЕНИ ОБРАБОТКИ ЗАПРОСОВ В ЗАДАЧЕ МОНИТОРИНГА ИНФОРМАЦИОННЫХ СИСТЕМ

Максимов А.Ю.

ФГБОУ ВО «МИРЭА – Российский технологический университет», Москва, e-mail: apcel@yandex.ru

Аннотация. В процессе эксплуатации информационных систем осуществляется мониторинг, позволяющий обеспечить их работоспособность и бесперебойную работу. При этом зачастую пороговые значения систем мониторинга задаются экспертами и не позволяют учитывать режим работы системы. В статье формулируются общие определения систем мониторинга, рассматривается подход к классификации данных мониторинга информационных систем на основе анализа распределения времени обработки http-запросов, проходящих через балансировщик нагрузки. Выделяются интервалы с максимальной плотностью распределения времени обработки запросов, описывается их физический смысл, на основе распределения составляется конечная цепь Маркова. Создается визуализация распределения запросов по уровням в зависимости от времени суток. На основе данных журналов балансировки за другой период времени делается вывод о необходимости повторного вычисления параметров метода при изменении конфигурации системы; на основе случайных выборок проверяется стабильность работы метода и информационной системы. Формулируется гипотеза о связи изменения характера распределения запросов по интервалам с изменением режима работы системы. Формулируется алгоритм применения данного метода на практике, его предположительные достоинства и недостатки, ограничения, выделяются дальнейшие направления исследований.

Ключевые слова: мониторинг информационных систем, балансировка нагрузки, цепи Маркова, анализ данных, время обработки http-запросов

MULTI-CRITERIA CLASSIFICATION OF INFORMATION SYSTEM MONITORING DATA

Maksimov A.Yu.

MIREA State University, Moscow, e-mail: apcel@yandex.ru

Annotation. During the operation of information systems, monitoring is used, which ensures its operability and uninterrupted operation. At the same time, the threshold values of monitoring systems are often set by experts and do not allow taking into account the operating mode of the system. The article formulates general definitions of monitoring systems, considers an approach to classification of information systems monitoring data based on an analysis of the distribution of processing time for http requests passing through a load balancer. Intervals are identified, in which the density of distribution of the request processing time is maximum, their physical meaning is described, a Markov chain is described based on the distribution. A visualization of the distribution of requests by level depending on the time of day is created. Based on the balancing log data for another period of time, it is concluded that the method parameters need to be recalculated when the system configuration is changed; on the basis of random samples, the stability of the method and the information system is checked. A hypothesis is formulated about the relationship between the change in the nature of the distribution of requests over intervals and the change in the operating mode of the system. An algorithm for applying this method in practice is formulated, its supposed advantages and disadvantages, limitations, and further areas of research are highlighted.

Keywords: information technology system monitoring, load balancing, Markov chains, data analysis, http requests processing time

Эксплуатация и сопровождение информационной системы (ИС) являются основным этапом жизненного цикла ИС, в интересах которого производится проектирование, разработка и тестирование программного обеспечения. С целью оперативного исправления возникающих проблем в работе ИС, их ретроспективного анализа используется мониторинг в виде автоматизированного наблюдения за соответствующим программным обеспечением. Однако зачастую пороговые значения системы мониторинга задаются на основе экспертных оценок, что не всегда позволяет отслеживать изменение параметров работы системы. В статье предлагается подход к выделению профиля работы распределенной информационной системы на основе анализа времени обработки запросов на примере http-запросов.

Цель исследования – рассмотреть подход к выделению пороговых данных системы мониторинга, позволяющий снизить влияние экспертных оценок на процесс мониторинга ИС.

Материалы и методы исследования

К числу основных понятий мониторинга ИС относятся метрики и триггеры. Ме-

трики – элементы данных, сохраняющие текущее и историческое значение конкретного параметра эксплуатируемой системы. Они могут принимать числовые значения, хранить текстовые данные, агрегироваться системой мониторинга по заданному правилу, либо генерироваться и собираться эксплуатируемой системой мониторинга при помощи механизма автоматического обнаружения метрик. Таких метрик может быть огромное множество – как стандартных, отражающих потребление аппаратных ресурсов USE-метрик [1], так и специфичных, например, бизнес-метрик, отражающих параметры протекающих в системе бизнес-процессов [2].

Триггеры – правила или пороговые значения конкретной метрики, набора метрик. При соблюдении заданного правила создается событие мониторинга. Событие мониторинга может отображаться по запросу или автоматически направляться адресатам в виде информационного сообщения; система мониторинга может выполнять заданные действия при возникновении события. Правила обработки метрик зачастую задаются вручную и имеют в своей основе нечеткую логику: «если значение количественного параметра А > 0 и значение количественного параметра Б = 0, сгенерировать событие». Некоторые системы мониторинга позволяют совершать операции над историческими данными: «если максимальное значение количественного параметра А последние 10 минут не превышает 15, сгенерировать событие мониторинга». Подобные правила зачастую задаются вручную на основе интуитивных догадок управляющих ИС мониторинга экспертов.

Визуализация – это отображение выбранных элементов данных и их исторических значений на графиках и их коллекциях. Она позволяет визуально анализировать тренды изменения параметров работы ИС, определять пороговые значения для задания триггеров или исправлять не покрытые правилами создания событий ситуации, заводить новые триггеры.

Таким образом, можно констатировать, что распространен анализ работы ИС на основе мониторинга ее ПО вручную, по интуитивно понятным алгоритмам и методам, что ведет к низкой оперативности обработки результатов мониторинга и их недостаточной достоверности в выявлении причин изменения показателей (характеристик) работы ИС.

В настоящей статье предлагается формальный метод построения триггеров для метрик, измеренных в количественной шкале. В качестве примера использован класс метрик, характеризующих выход режима работы ИС за стандартные рамки, причиной чего могут выступать сбой оборудования, программ, отклонение соответствующих показателей от их ожидаемых значений. Для определенности рассмотрим одну из таких метрик – время обслуживания http-запросов, проксируемых на одном централизованном балансировщике нагрузки. Наличие централизованного балансировщика позволяет обрабатывать данные в едином узле, избегая связанных с согласованностью данных проблем [3] и упрощая алгоритм работы мониторинга.

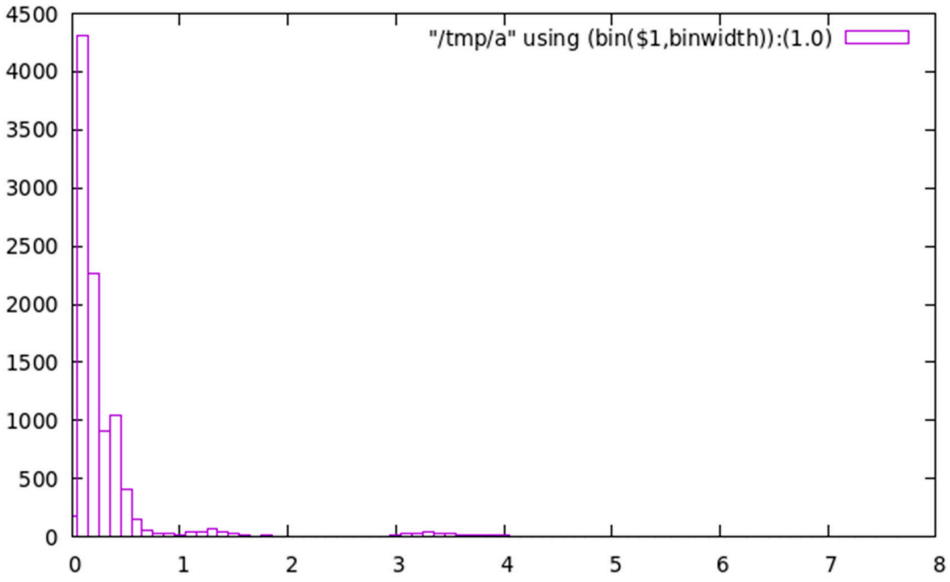

Рис. 1. Гистограмма распределения времени обслуживания http-запросов: по горизонтали отражено время обработки запроса, по вертикали – количество подобных запросов в выборке

На рис. 1 представлена гистограмма распределения времени обслуживания запросов, составленная с точностью в 0,1 с.

В данных времени обработки http- запросов выделены критические уровни (КУ), определяемые значениями времен обслуживания, в небольших интервалах которых сосредоточено наибольшее количество значений (с фиксированными временными характеристиками http-запросов). Предполагается, что изменение характера распределения временных параметров характеризует качественные изменения в процессах исследуемой стабильно работающей ИС. Поэтому определение КУ оценивается как актуальная задача, решение которой позволит повысить эффективность мониторинга ИС.

Для понимания физического смысла критических уровней с точки зрения задач мониторинга ПО остановимся на полученных результатах. Получены 4 интервала значений. Для удобства изложения присвоим им идентификаторы от ‘0’ до ‘3’ со следующей характеризацией:

− Интервал «0»: от 0 с до первой критической точки. Быстрые обращения в кэш, запись xml-сообщений без сертификатов ЭП.

− Интервал «1»: между 1 и 2 критическими точками. Обращения без кэширования, с записью одного небольшого объекта.

− Интервал «2»: между 2 и 3 критическими точками. Обращения с записью сообщений, содержащих несколько небольших объектов.

− Интервал «3»: после 3 критической точки. Обращения с записью больших объектов, обращения с записью множества различных объектов, прочие занявшие продолжительное время.

Построен график распределения количества обращений по времени с точностью в 1 ч. Результаты приведены на рис. 2.

Четко видно суточную составляющую характера данных, присущую данной системе. Распределение типов запросов в целом отражает ожидаемый результат в течение дня и возникает из-за суточного характера поступающих в информационную систему заявок, особенностей их обработки.

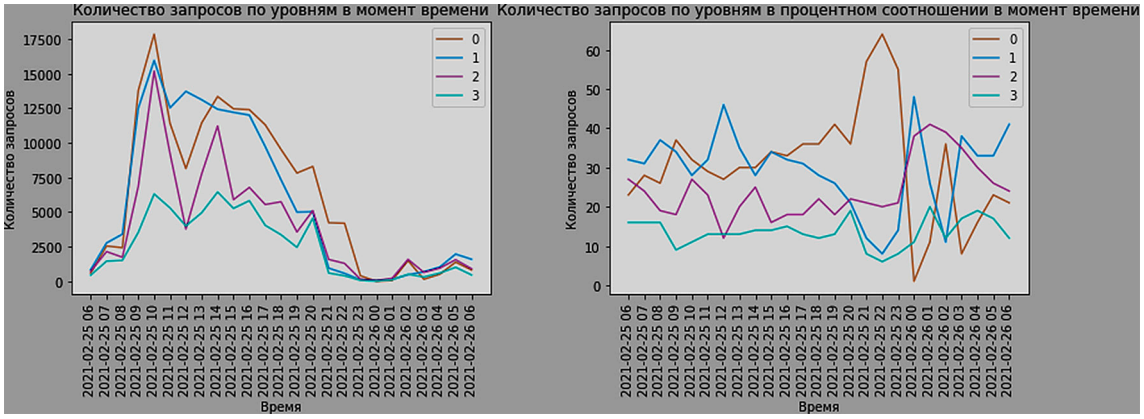

Рис. 2. Распределение типов обрабатываемых запросов по времени

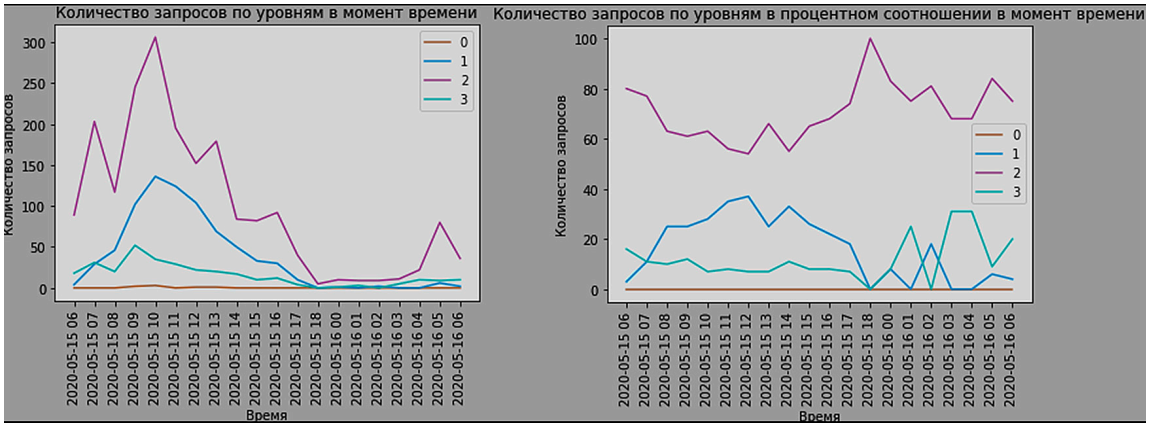

Рис. 3. Применение информации о критических уровнях в системе другой конфигурации

Матрица вероятности совершения запроса одного типа после другого

	0	1	2	3
0	0.48	0.28	0.13	0.09
1	0.29	0.38	0.19	0.11
2	0.22	0.26	0.36	0.13
3	0.24	0.27	0.2	0.28

Также полученные уровни применены при обработке данных этой же системы в другой конфигурации в другое время. Характер распределения данных сильно отличается, в качестве причины выступает сильное изменение конфигурации оборудования и программного обеспечения. Результаты приведены на рис. 3.

Подобные результаты могут свидетельствовать о необходимости повторного расчета уровней при сильном изменении конфигурации системы; однако при отсутствии изменений в конфигурации это может свидетельствовать об изменении характера работы системы и необходимости проведения анализа причин возникшей ситуации.

Далее на основе данных о распределении построена матрица вероятности совершения запроса каждого из типов после текущего запроса. Здесь строки – конкретные запросы; ячейки в строках – вероятности совершения запроса следующего типа.

Из таблицы можно сделать вывод, что после совершения каждого запроса наиболее вероятно совершение запроса такого же типа; однако для «долгих» обращений эта вероятность не так выражена.

Предполагается, что в дальнейшем полученные результаты позволят повысить информативность системы мониторинга для данной системы и обращающихся к ним систем: к примеру, сильное преобладание в оперативных данных запросов того или иного типа отразится на матрице вероятности переходов между типами запросов и их количественном и процентном соотношении, что может сигнализировать о не обнаруживаемом прочими средствами мониторинга программном либо аппаратном сбое. Были высчитаны аналогичные матрицы по отрезкам исторических данных меньшего размера: распределение значений в ячейках незначительно отличается от рассчитанных по полным данным, что свидетельствует о стабильности как метода, так и работы системы.

В качестве дальнейшего направления развития работы можно выделить проверку предположения о связи сбоев и распределения запросов по уровням на реальных исторических данных комплекса мониторинга данной информационной системы, применение метода к данным других модулей информационной системы.

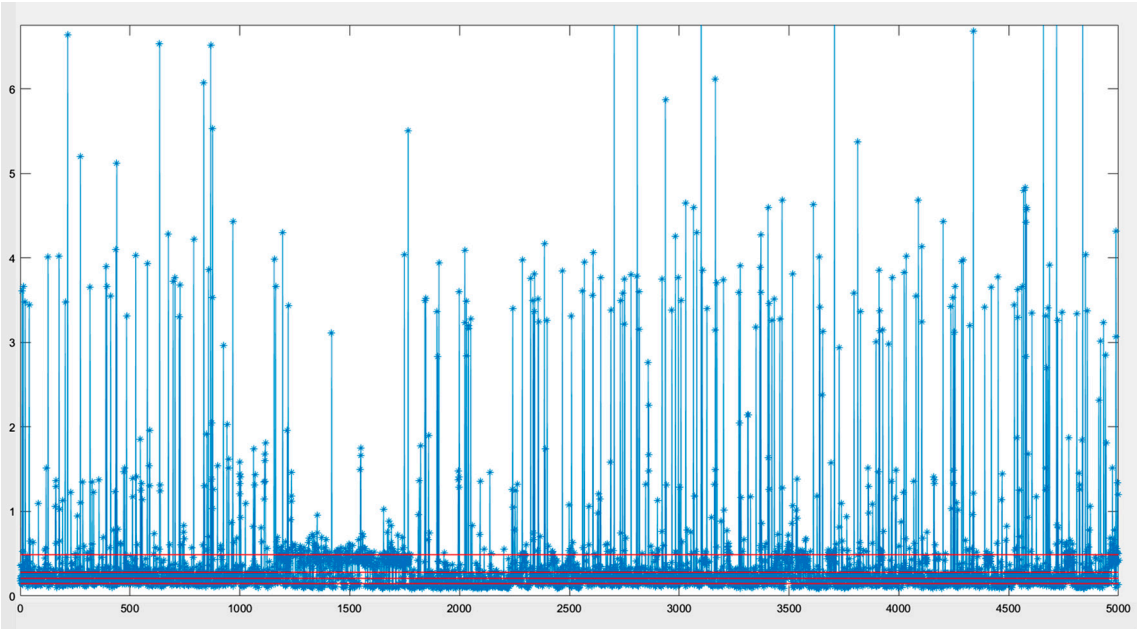

Рис. 4. Аппроксимация последовательности времен совершения запросов марковской цепью

Предложенный подход к формализации последовательности запросов в виде марковской цепи позволяет прогнозировать появление сложных запросов, требующих использования больших ресурсов и длительных временных затрат в их обслуживании, а также отслеживать расхождение прогноза поведения информационной системы с реальным поведением. Это может значительно повысить эффективность проведения мониторинга ПО ИС. Пример подобной аппроксимации приведен на рис. 4.

Суть аппроксимации временной последовательности конечной цепью Маркова (КЦМ) заключается в ее многоуровневой градации таким образом, что каждый из уровней разбивает график на полосы, границы которых определяют переходы последовательности точек КЦМ в соседние состояния. Такое геометрическое представление позволило разработать простой алгоритм определения матрицы вероятностей указанной КЦМ, полностью согласующийся с методом, изложенным в [4, с. 159–169]. Указанный в данном источнике метод опубликован в 1970 г. – однако он относится к фундаментальной теории КЦМ, поэтому актуален на момент написания текста.

Для оценки временных характеристик достижения критических уровней можно использовать подход [5, с. 152–159], суть которого заключается в построении соответствующей КЦМ поглощающей цепи Маркова с матрицей вероятностей, представленной в канонической форме, удобной для расчета указанных временных характеристик применительно к задаче мониторинга ПО ИС.

Результаты исследования и их обсуждение

Как было упомянуто ранее, практическое применение метода может выражаться следующим образом:

1. На основе исторических данных журнала балансировки нагрузки рассчитываются КУ и параметры КЦМ, матрица вероятности совершения переходов между звеньями КЦМ, или вероятности совершения попадающего в один критический уровень запроса после предыдущего.

2. Данные апробируются на продолжительном интервале времени, различных стендах с имитирующей продуктивную нагрузкой. Для стендов с различными характеристиками аппаратного и программного обеспечения (АО и ПО) определяются отличные от исходных КУ – это позволяет использовать одну матрицу КЦМ для сравнения различных стендов при сохранении характера нагрузки. По необходимости – к примеру, при выявлении наличия значимых для весов матрицы сбоев на проанализированных временных отрезках – заново высчитываются КУ и/или матрица КЦМ.

3. В процессе мониторинга ИС непрерывно-кусочно высчитывается значение вероятности совершения запросов одного типа после запросов другого. При выявлении значимых отличий между текущей и характерной для данного режима работы матриц генерируется событие мониторинга, свидетельствующее о возможном выходе режима работы системы за стандартные рамки.

4. При изменении характеристик программно-аппаратного обеспечения проверяется необходимость корректировки исходных данных метода; при необходимости выполняется пересчет значений.

К предположительным достоинствам данного метода можно отнести:

− Оценку характеристики функционирования системы в конкретный момент времени, ее отличие от среднего профиля.

− Возможность анализа схожести различных стендов с промышленным.

− Автоматизированное вычисление характеристик работы системы на основе результатов ее работы.

− Простоту работы метода: при анализе данных используется только один параметр, время совершения запроса, и численные операции с плавающей запятой, что вносит низкую дополнительную нагрузку в работу ИС мониторинга и практически не требует дополнительных вычислительных ресурсов от стенда и ИС мониторинга.

К предположительным недостаткам данного метода можно отнести:

− Необходимость работы системы под нагрузкой: к примеру, в проанализированном интервале времени среднее количество совершаемых запросов держится на уровне 10 запросов в секунду.

− Стабильность работы системы: при невозможности построения матрицы вероятности совершения запросов одного после другого данный метод неприменим.

− Наличие в совершаемых запросах различного физического смысла: при обработке сервисом только однообразных запросов данный метод неприменим.

− Необходимость накопления статистики и повторного вычисления параметров после значимых изменений АО и ПО.

− Возможную неприменимость метода из-за индивидуальных особенностей информационной системы.

Заключение

В данной работе приведены общие сведения о подходах к мониторингу информационных систем, был исследован подход к мониторингу на основе классификации поступающих запросов по времени ответа. Данный подход позволяет выявить закономерности в работе информационной системы и выявлять отклонение поведения системы от обычного режима работы, снижая влияние экспертной оценки на подсистему мониторинга.

Развитие предложенного метода может быть продолжено в последующих статьях в части расширения классификации запросов, учета других сохраняющихся в журналы балансировки нагрузки параметров.

Список литературы

1. Захаров И.Е., Панарин О.А., Рыкованов С.Г., Загидуллин Р.Р., Малютин А.К., Шкандыбин Ю.Н., Ермекова А.Е. Мониторинг приложений на кластере ZHORES в Сколтехе // Программные системы: теория и приложения. 2021. Т. 12, № 2 (49). С. 73–103.

2. Альфара А.А.Ю., Королев Д.В., Зайцев К.С., Дунаев М.Е. Разработка системы мониторинга для серверного приложения // International Journal of Open Information Technologies. 2023. Т. 11, № 8. С. 24–31.

3. Сычугов А.А., Греков М.М. Централизованно-распределенная модель системы обнаружения аномалий // Известия Тульского государственного университета. Технические науки. 2021. №. 5. С. 316–322.

4. Кемени Д.Д., Лори Снел Дж. Конечные цепи Маркова / пер. с англ. С.А. Молчанова и др.; под ред. А.А. Юшкевича. М.: Наука, 1970. 272 с.

5. Кельберт М.Я., Сухов Ю.М. Вероятность и статистика в примерах и задачах. М: МЦНМО, 2018. 486 с.