

# РАЗРАБОТКА ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ И МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ОПТИМИЗАЦИИ МОНИТОРИНГА

Шишкин Ю.Е. Email: Shishkin1133@scientifictext.ru

Шишкин Юрий Евгеньевич – аспирант,  
ORCID: 0000-0001-7442-2722,  
кафедра информационных технологий и компьютерных систем,  
Севастопольский государственный университет, г. Севастополь

**Аннотация:** в статье анализируются математическая модель и реализующая ее программа, предназначенные для снижения объема данных мониторинга в сложных, закрытых, недетерминированных многопараметрических системах. Метод решения основан на применении гибких решающих правил, использующих систему статистических метрик и дискретно-событийной имитационной модели. Область применения модели – оптимизация мониторинга в облачных сервисах как компонента обеспечения идеологии интернета вещей, автоматизации процессов в различных сферах деятельности и исключения из них человека.

**Ключевые слова:** Большие Данные, оптимизация мониторинга, система поддержки принятия решений, интернет вещей, система массового обслуживания, имитационное моделирование, визуализация.

## DEVELOPMENT OF TOOLS AND MATHEMATICAL MODELS FOR MONITORING OPTIMIZATION

Shishkin Yu.E.

Shishkin Yuriy Evgenevich – postgraduate,  
ORCID: 0000-0001-7442-2722,  
INFORMATION TECHNOLOGY AND COMPUTER SYSTEMS DEPARTMENT,  
SEVASTOPOL STATE UNIVERSITY, SEVASTOPOL

**Abstract:** the article analyzes the mathematical model and implementing a model program, designed to reduce the amount of monitoring data in complex, closed, non-deterministic multivariable systems. The solution method is based on the use of flexible decision rules, using a statistical metrics system and discrete event simulation model. Application models is cloud services monitoring optimization as an Internet of Things ideology software component, automation process in various fields of human activity to exclude a human from that fields.

**Keywords:** Big Data, monitoring optimization, decision support system, Internet of Things, queuing system, simulation, visualization.

УДК 004.6:004.415.2

### Введение

На сегодняшний день в связи с активным внедрением идеологии Больших Данных, интернета вещей и облачных вычислений значительно возросли требования к производительности, точности и оптимизации методов и средств мониторинга [1 - 3]. Известно, что при продолжительном поступлении практически идентичных сообщений, содержащих значения параметров объекта мониторинга, приращение количества информации на каждое следующее сообщение будет асимптотически стремиться к нулю, а значение информационной энтропии стремиться к некоторой константе [4 - 5].

Данный факт можно легко доказать исходя из определения собственной информации следующим образом:

$$I(X) = -\log P_X(X), \quad (1)$$

при этом, если значение параметра объекта мониторинга не изменяется или его изменения закономерны  $p(x_{t+1} | x_t) \rightarrow 1$ , что при известном  $x_t$  можно понимать как  $p(x_{t+1}) \rightarrow 1$ , получим  $I(x_{t+1}) \rightarrow 0$ , что и требовалось доказать.

Из чего следует, что разработка эффективных метрик, численно определяющих значимые изменения наблюдаемого параметра, позволит отфильтровать незначимые записи, что в свою очередь уменьшит нагрузку на систему хранения данных мониторинга и вычислительную сложность процесса предварительной обработки, кластеризации, визуализации и принятия решений [6-7]. Выделение из входящего потока данных только значений, соответствующих значимым изменениям в наблюдаемой системе позволит при неизменном объеме имеющихся в распоряжении вычислительных ресурсов увеличить число наблюдаемых объектов мониторинга или их параметров с целью увеличения производительности.

Целью исследования является разработка метода снижения объема данных мониторинга сложной, закрытой, недетерминированной многопараметрической системы и реализующей его программы. Для достижения поставленной цели необходимо решить следующие задачи:

- выработать систему метрик, определяющих меру значимости, поступающих в систему мониторинга данных;
- определить границы решающих правил о признании поступающих в систему мониторинга данных – незначимыми;
- построить имитационную модель, реализующую решающие правила и имитирующую процесс мониторинга параметров объекта.

Детектирующая метрика возмущения в поступающем потоке данных мониторинга, определяется как факт выхода, поступающего во входном потоке данных, значения за регламентированные пределы [8-9]. Расчет пределов происходит по данным, попавшим в пространственно-временное окно в соответствии с формулами:

$$Q_1 = x_{Q_1}^{\min} + \frac{i}{f_{Q_1}} \cdot \left( \frac{\sum_{i=1}^n f_i}{\sum_j x_j} - S_{Q_1-1} \right), \quad (2)$$

$$Q_2 = \frac{x_N + x_{N+1}}{2}, \quad (3)$$

$$Q_3 = x_{Q_3}^{\min} + \frac{i}{f_{Q_3}} \cdot \left( \frac{3 \cdot \sum_{i=1}^n f_i}{\sum_j x_j} - S_{Q_3-1} \right), \quad (4)$$

где  $x_{Q_1}^{\min}$  – нижняя граница первого квартиля;  $x_{Q_3}^{\min}$  – нижняя граница третьего квартиля;  $\sum_{i=1}^n f_i$  – число наблюдений;  $S_{Q_1-1}$  – значение плотности вероятности в конце диапазона первого;  $S_{Q_3-1}$  – значение плотности вероятности в конце третьего интервала;  $f_{Q_1}$  – частота поступления данных в первом интервале;  $f_{Q_3}$  – частота поступления данных в третьем интервале;  $N$  – количество значений в совокупности данных;  $i$  – величина интервала,  $\sum_j x_j$  – число интервалов.

### Организация и проведение экспериментов

Исходные параметры моделирования были определены в панели свойств генератора в системе AnyLogic, как показано на рисунке 1, следующим образом:

- время между поступлениями заявок входного потока имеет экспоненциальное распределение с параметром  $\lambda = 1$ ;
- каждому агенту при выходе, параметру monitoredValue присваивается случайное значение, имеющее нормальное распределение с коэффициентом сдвига  $\mu = 5$  устанавливаемым значением ползункового манипулятора и коэффициентом масштаба  $\sigma = 5$ .
- активным является только первый генератор агентов, для остальных генераторов значение интенсивности устанавливается равным нулю, что можно интерпретировать как мониторинг только одного параметра.

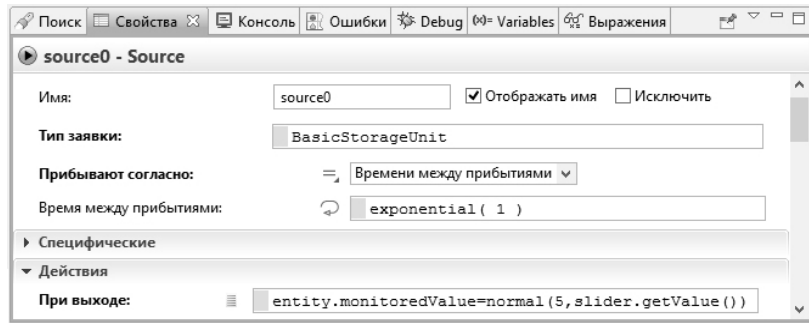


Рис. 1. Параметры стандартного генератора заявок в системе AnyLogic

При выходе агента из очереди и поступлении на компонент диспетчер заявок производится вызов функции `decisionFunction`. Параметры диспетчера показаны на рисунке 2.

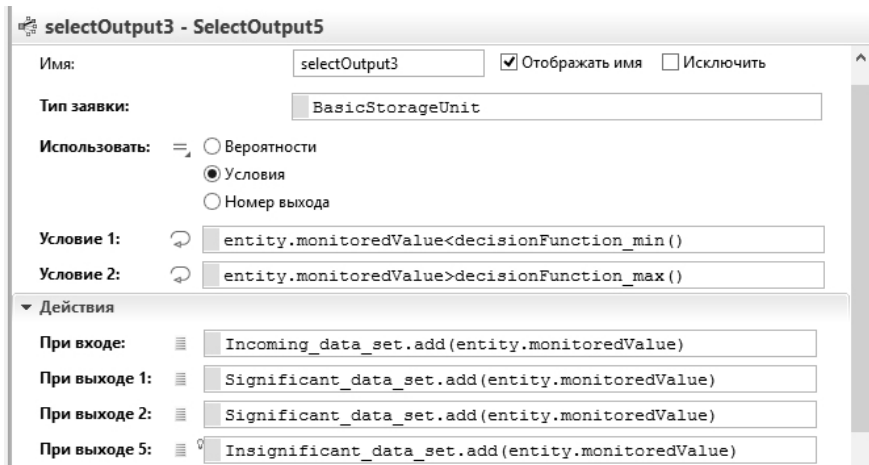


Рис. 2. Параметры стандартного диспетчера заявок в системе AnyLogic

Функции `decisionFunction_min()` и `decisionFunction_max()` определяют, входит ли значение наблюдаемого параметра в интервалы, ограниченные 1 и 4 квартилем, если входит, то данные считаются значимыми, заносятся в «Significant data set» и отображаются на соответствующей гистограмме, иначе – игнорируются, но в программной реализации модели заносятся в «Insignificant data set» и отображаются на соответствующей гистограмме.

Произведем запуск модели с указанными значениями, начало процесса выполнения представлено на рисунке 3. На момент запуска модели размер контрольной выборки, содержащейся в «Incoming data set» – нулевой, первые три поступивших значения были признаны значимыми, а четвертое значение имело значение близкое к математическому ожиданию контрольной выборки и записано в «Insignificant data set».

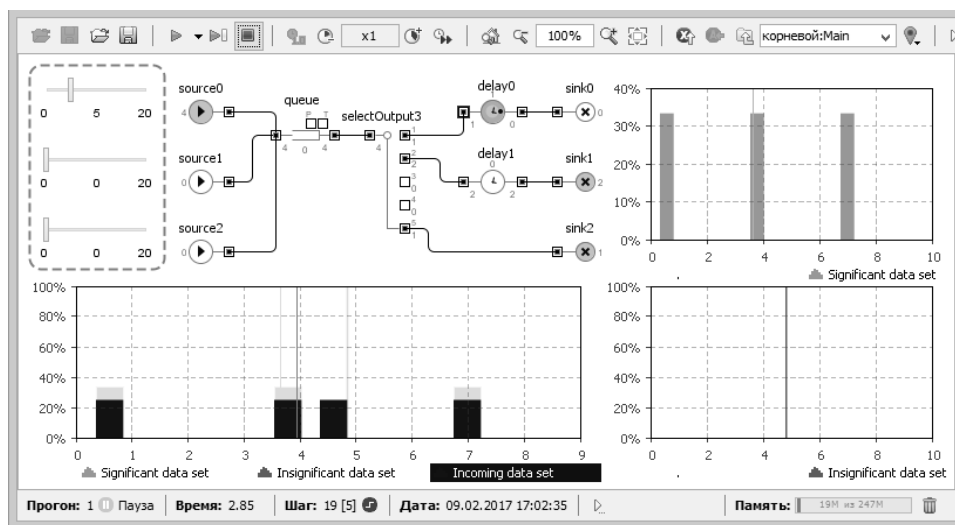


Рис. 3. Графический интерфейс состояния модели в начале моделирования

После генерации первых 50 агентов, как показано на рисунке 4, становится визуально заметным закон распределения наблюдаемой случайной величины по гистограмме «Incoming data set».

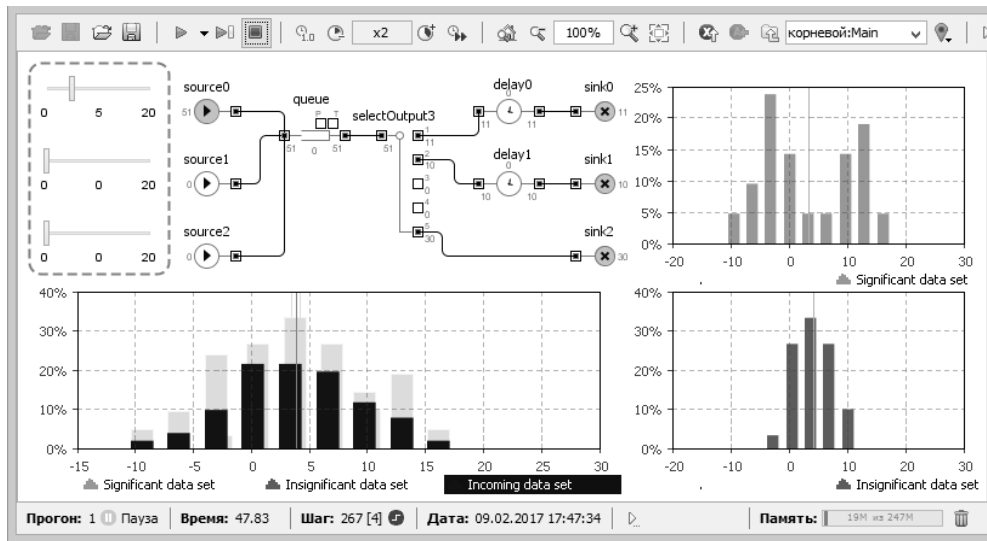


Рис. 4. Графический интерфейс состояния модели на 50 шаге моделирования

Вследствие того, что все гистограммы имеют одинаковые граничные интервалы можно визуально судить о характере распределения данных в множества значимых – «Significant data set» и незначимых «Insignificant data set» данных. Если суммировать данные обеих гистограмм, то получим исходную «Incoming data set». Исходя из определения критерия квантификации, применительно к данной системе, принятым информационным фильтром должно отсеиваться в среднем 50% поступающих данных, при условии отсутствия выбросов и работы в установившемся стационарном режиме. В рассматриваемом случае 20 записей были признаны значимыми, а 30 – незначимыми, таким образом, 60% данных признаны системой незначимыми, что обусловлено стабильностью значений наблюдаемой случайной величины и недостаточным объемом контрольной выборки в момент запуска модели.

Продолжим моделирование системы с неизменными параметрами в течение 500 модельных временных единиц, результат представим на рисунке 5.

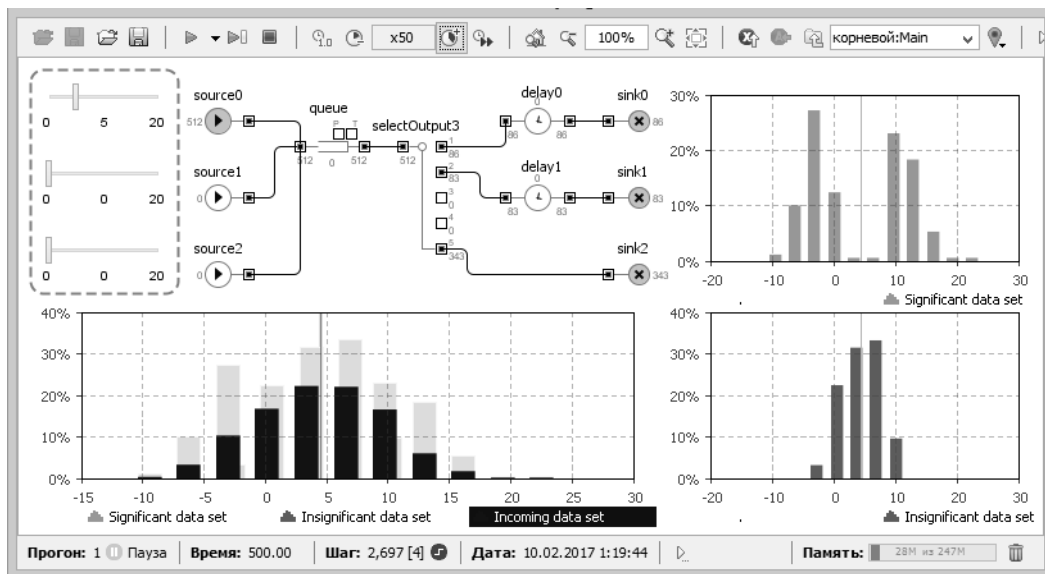


Рис. 5. Графический интерфейс состояния модели на 500 шаге моделирования

Результат работы системы на 500 шаге, в установившемся режиме, становится более ярко выраженным по сравнению с результатом на 50 шаге, особенно явно это проявляется по провалам плотности распределения записей, в окрестностях математического ожидания на гистограмме «Significant data set», незначимыми были признаны 67% наблюдений.

Представляет интерес процесс реакции системы на постоянное изменение значения функции сдвига наблюдаемого параметра [10 - 12]. В моделируемом случае объем «Incoming data set» будем считать

достаточно большим, что должно привести к достаточно высокой инертности системы, более высокой чувствительности к единичным выбросам и нетолерантности, в том числе, к постоянным изменениям математического ожидания, увеличению дисперсии и изменению функции распределения наблюдаемой случайной величины.

Примером применения таких параметров эксперимента может служить модель мониторинга температуры пациента, любое изменение температуры за пределы нормального диапазона будет считаться значимым, при этом, если повышенная температура будет наблюдаться постоянно, то значимым будет считаться даже малое ее увеличение, но стабильно повышенная температура не будет приводить к появлению сигнала тревоги, другими словами происходит адаптивное повышение границы чувствительности наблюдаемого параметра при его отклонении от контрольной выборки, при этом допускается незначительный дрейф границ допустимого интервала.

Произведем двукратное увеличение значения параметра сдвига в нормальном распределении наблюдаемой случайной величины с 5 до 10, результаты представим на рисунке 6.

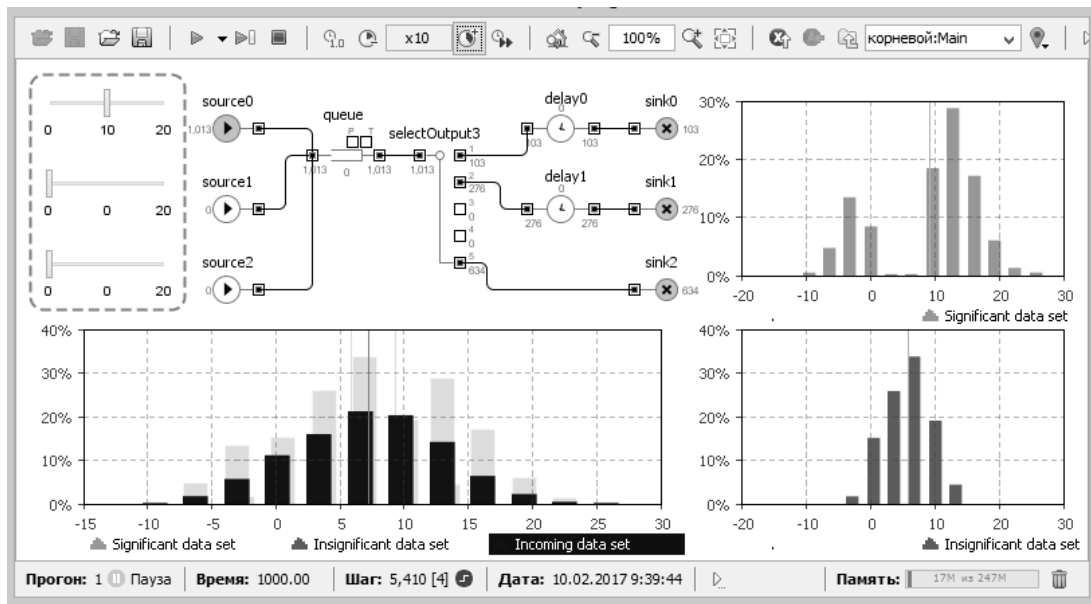


Рис. 6. Графический интерфейс состояния модели на 1000 шаге моделирования при увеличенном параметре сдвига наблюдаемой случайной величины

В результате увеличения параметра сдвига с 5 до 10 при нормальном распределении происходит соответственное увеличение математического ожидания с 5 до 10. При этом в период с 500 по 625 шаг система принятия решений сигнализировала о появлении значимых изменений, и наблюдалось значительное увеличение объема матрицы значимых наблюдений «Significant data set». В период с 650 по 1000 шаг за счет расширения контрольной выборки система обучилась, и дальнейшее появление агентов с новым значением математического ожидания не приводило к появлению сигнала тревоги и отмечалось как незначимое, незначимыми были признаны 62,5% наблюдений.

### Заключение

В отличие от задач, решаемых в [10 - 12], где рассматривалась статическая модель среды, в которой априорно известно, что найденные оптимальные параметры информационных метрик могут быть использованы для всех стационарных состояний объекта управления, в настоящей работе учитываются неопределенности внутренних состояний объекта мониторинга, вплоть до появления новых состояний, проявляющихся в изменении математического ожидания, дисперсии и закона распределения наблюдаемых параметров.

К достоинствам разработанной программной модели можно отнести возможность переобучения системы принятия решений о наступлении значимых изменений в динамике наблюдаемого параметра в зависимости от закона распределения наблюдаемой случайной величины, возможность динамического уменьшения порога срабатывания детектора при продолжительном пребывании системы в стабильном состоянии.

В соответствии с экспериментальными данными, полученными в результате имитационного моделирования, для верхней и нижней границы решающего правила о признании данных незначимыми  $x_{Q_1}^{\min} = 0,25$  и  $x_{Q_3}^{\max} = 0,75$  соответственно, позволили отсеять 75% незначимых наблюдений.

1. *Пасынков М.А.* Комплексная система интеграции баз данных мониторинга физических параметров и позиционирования в акваториях // Научный журнал, 2017. № 2 (15). С. 29 - 31.
2. *Греков А.Н., Шишкин Ю.Е.* Моделирование трехкомпонентного акустического измерителя скорости течения // Системы контроля окружающей среды. Севастополь: ИПТС, 2016. № 6 (26). С. 33-40.
3. *Шишкин Ю.Е.* Облачные сервисы в системах поддержки принятия решений // Научный журнал, 2017. № 1 (14). С. 19 - 20.
4. *Shishkin Y.E.* Big Data visualization in decision making // Science in Progress, 2016. С. 203-205. ISBN 978-5-7782-3094-1.
5. *Шишкин Ю.Е.* Визуальный анализ Больших Данных с применением познавательных паттернов // Проблемы современной науки и образования, 2017. № 2 (84). С. 24 - 26.
6. *Шишкин Ю.Е., Греков Н.А.* Исследование систем управления высокоточными измерениями // Интеллектуальные системы, управление и мехатроника, 2015. С. 221 - 225.
7. *Кодолов П.А.* Облачное хранилище данных // Наука, техника и образование, 2016. 2016 № 4 (22). С. 51-53.
8. *Шишкин Ю.Е.* Анализ моделей взаимодействия пользователей и провайдеров облачных сервисов // Интеллектуальные системы, управление и мехатроника, 2016. 2016. С. 289 - 293.
9. *Шишкин Ю.Е.* Актуализация данных в системах мониторинга сложных объектов с использованием информационных метрик // Проблемы современной науки и образования, 2017. № 6 (88). С. 22 - 27. doi:10.20861/2304-2338-2017-88-001.
10. *Скатков А.В., Брюховецкий А.А., Шишкин Ю.Е.* Сравнительный анализ методов обнаружения изменений состояний сетевого трафика // Автоматизация и приборостроение: проблемы, решения, 2016. С. 14 - 15.
11. *Магжанова А.Т.* Применение облачных технологий для реализации решений интернета вещей // Современные инновации, 2016. № 7 (9). С. 30 - 34.
12. *Шишкин Ю.Е., Скатков А.В.* Решение задачи составления расписаний большой размерности с применением технологии Больших Данных // Информационные технологии и информационная безопасность в науке, технике и образовании. «ИНФОТЕХ – 2015», 2015. С. 103 - 105.