УДК 004.051 : 658.562.012.7

ИСПОЛЬЗОВАНИЕ КОНТРОЛЬНЫХ КАРТ ДЛЯ ОБНАРУЖЕНИЯ ФРОДА ПРИ ПЕРЕДАЧЕ ГОЛОСОВОГО ТРАФИКА

© 2024 А.А. Санталов[1], В.Н. Клячкин[2]

[1] ООО «Телесистемс», г. Санкт-Петербург, Россия
[2] Ульяновский государственный технический университет, г. Ульяновск, Россия

Статья поступила в редакцию 06.06.2024

Непрерывное развитие мошеннического телекоммуникационного трафика требует повышения качества работы систем борьбы со спамом и фродом – антифрод-систем. Модуль антифрод-системы по обнаружению аномального трафика использует для мониторинга контрольные карты, построенные по временным рядам характеристик процесса передачи голосового трафика, которые обладают нестационарностью, цикличностью и асимметричным распределением, что затрудняет их анализ. В статье представлен алгоритм, снижающий частоту ложных срабатываний контрольных карт и повышающий эффективность работы антифрод-системы. Алгоритм основан на подготовке исходного временного ряда характеристики процесса передачи голосового трафика к анализу с помощью контрольных карт путем сегментации нестационарного ряда с циклической составляющей и последующего объединения однородных сегментов в одну общую выборку.
Ключевые слова: контрольные карты индивидуальных наблюдений и скользящих размахов, сегментация временного ряда, нестационарный временной ряд, антифрод-система, мониторинг.

DOI: 10.37313/1990-5378-2024-26-4(3)-395-399
EDN: MXBPTB

ПОСТАНОВКА ЗАДАЧИ

Непрерывное развитие мошеннического телекоммуникационного трафика [1] приводит к росту убытков операторов связи. Это вызывает необходимость совершенствования антифрод-систем (систем борьбы со спамом и фродом — мошенническим трафиком) и повышения качества их работы. Модуль обнаружения антифрод-системы отвечает за выявление временных срезов аномального трафика, которые будут анализироваться на предмет наличия мошеннических звонков модулем анализа.

Для обнаружения аномального трафика применяются методы, использующие данные подробной записи о вызове (CDR — call detail record), содержащей информацию о времени начала звонка, длительности звонка, номере абонента А, номере абонента Б и т. д. Из набора CDR извлекается дополнительная информация путем агрегирования данных по маршрутам передачи звонков и временным интервалам для получения временных рядов характеристик процесса передачи голосового трафика. Маршрут с точки зрения транзитного оператора – это путь прохождения звонка, включающий в себя клиента-оператора, отправившего запрос на установление соединения, страну абонента А, телефонную сеть в стране абонента Б и поставщика-оператора, который это соединение устанавливает.

Традиционно в отрасли для мониторинга используются характеристики качества ACD (*average call duration* — средняя длительность звонков) и ASR (*answer seizure ratio* — коэффициент полученных ответов). Временные ряды этих характеристик обладают цикличностью, стационарностью и асимметрией распределения. В [2] предложена система методик статистического управления процессом, которая учитывает эти свойства временных рядов. ACD и ASR используются преимущественно при мониторинге качества связи, так как снижение этих характеристик сигнализирует об ухудшении связи, когда абоненты реже дозваниваются, а во время связи происходят обрывы. Эти характеристики могут применяться при обнаружении некоторых видов мошеннического трафика, однако мошеннический трафик смешан с трафиком абонентов связи и не является доминирующей частью трафика, поэтому из-за нормализующего свойства центральной предельной теоремы значения этих характеристик могут отклоняться незначительно.

Мошеннический трафик добавляется к трафику абонентов, поэтому его появление характеризуются всплеском (резким увеличением) количества звонков в час и количества минут

Санталов Антон Александрович, системный аналитик отдела статистических исследований и разработок, аспирант кафедры прикладной математики и информатики Ульяновского государственного технического университета (УлГТУ).
E-mail: anton.santalov1995@gmail.com
Клячкин Владимир Николаевич, доктор технических наук, профессор кафедры прикладной математики и информатики УлГТУ. E-mail: v_kl@mail.ru

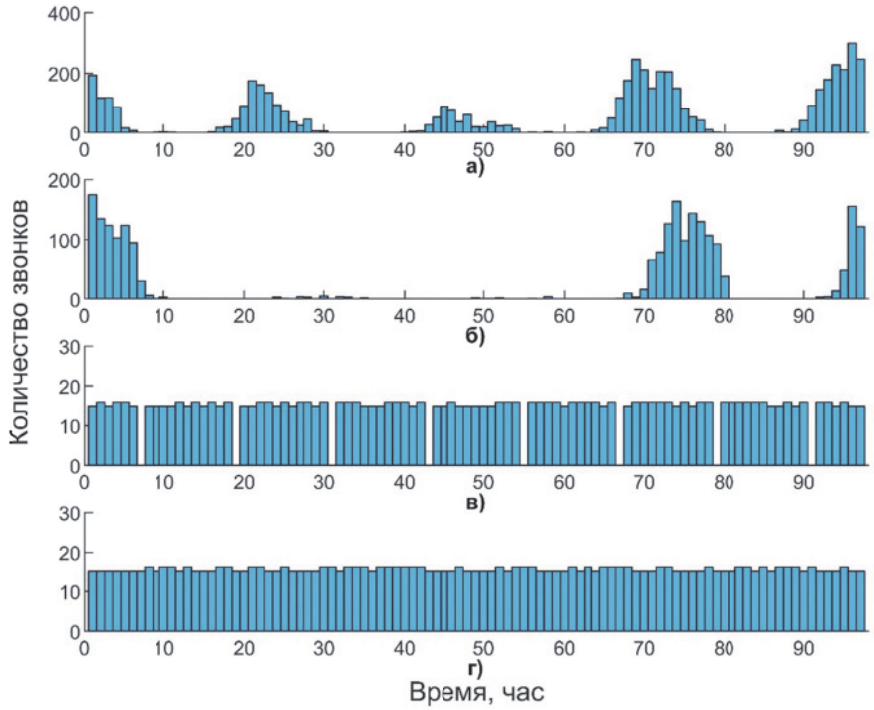

Рисунок 1. Временные ряды процесса передачи голосового трафика

трафика в час. По этой причине целесообразно использование именно этих характеристик для построения контрольных карт при мониторинге процесса передачи трафика на предмет появления мошенничества. Однако временные ряды этих характеристик обладают нестационарностью и так же, как и ACD и ASR, цикличностью и асимметричным распределением, что усложняет их использование для построения контрольных карт. На рисунке 1 изображены графики временного ряда характеристики процесса передачи голосового трафика. Из графиков видно, как меняется количество звонков в зависимости от времени суток и дней недели у разных маршрутов: а) с трафиком обычных абонентов, б) с трафиком компаний и колл-центров, в) и г) с трафиком автоматизированных систем.

Если не учитывать эти особенности временного ряда, то это приводит к частым ложным срабатываниям контрольных карт. Снижение частоты ложных срабатываний является одной из задач по улучшению эффективности работы сложной технической системы.

АЛГОРИТМ ПОСТРОЕНИЯ КОНТРОЛЬНЫХ КАРТ

Для выявления мошеннического трафика использовано построение контрольных карт индивидуальных наблюдений и скользящих размахов [3-5] характеристики процесса передачи голосового трафика – количества звонков в час, передаваемых по маршруту. Предлагается алгоритм обнаружения нарушения процесса, включающий в себя подготовку данных путем разбиения временного ряда характеристики на сегменты в точках, где происходит изменение его характера, и устранение циклической составляющей с последующим объединением одинаковых по вероятностным характеристикам сегментов в выборку, по которой строятся контрольные карты. Алгоритм представлен на рисунке 2, жирными рамками выделены этапы, улучшающие алгоритм обнаружения нарушения процесса, использовавшийся в антифрод-системе изначально. Такая подготовка данных направлена на уменьшение количества ложных срабатываний модуля обнаружения на временных рядах вида а) и б) с рисунка 1.

На этапе 1 производится сегментация временного ряда. Модуль обнаружения подвергается высокой нагрузке, так как в режиме реального времени рассчитывает множество характеристик среза голосового трафика и проверяет срабатывание множества правил на каждом срезе трафика. По этой причине выбор может быть остановлен на методе сегментации, который не требует много вычислительных ресурсов. В качестве таких методов можно использовать PELT (англ. Pruned Exact Linear Time – сокращенный точный метод сегментации, работающий за линейное время), байесовский метод обнаружения точек изменения (англ. Bayesian Changepoint Detection), бинарная сегментация (англ. Binary Segmentation), современные методы ClaSP (англ. Classification Score Profile – профиль классификационных оценок) или ClaSS (англ. Classification Score Stream – поток классификационных оценок) [6-8].

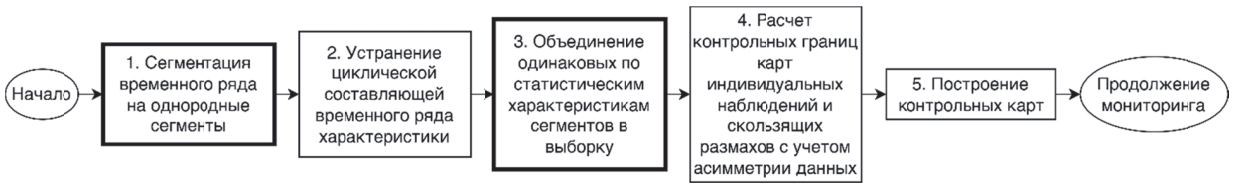

Рисунок 2. Алгоритм обнаружения нарушения процесса с помощью контрольной карты

На этапе 2 происходит исключение циклической составляющей временного ряда [9, 10].

На этапе 3 попарно сравниваются сегменты временного ряда с последним его сегментом. Чтобы объединить выборки двух сегментов вместе, необходимо убедиться в их принадлежности одной генеральной совокупности. Можно использовать критерий Шапиро-Уилка для проверки нормальности распределения двух выборок. Если обе сравниваемые выборки имеют нормальное распределение, то для проверки гипотезы о равенстве дисперсий двух выборок применяется критерий Фишера. Если хотя бы одна из сравниваемых выборок не имеет нормальное распределение, то над выборками производится преобразование Бокса-Кокса [11]. Затем проверяется гипотеза о равенстве средних значений двух выборок. Если эта гипотеза принимается, то сегмент добавляется к последнему сегменту временного ряда в общую выборку, которая будет использована для построения контрольных карт.

После формирования общей выборки на этапе 4 вычисляются контрольные границы контрольных карт индивидуальных наблюдений и скользящих размахов с учетом асимметрии распределения значений [12].

На этапе 5 строятся контрольные карты и проверяется выход точек за контрольные границы. Выход точки за границу карты индивидуальных наблюдений даст сигнал о том, что количество звонков за текущий час нехарактерно для данного маршрута. Выход точки за границы карты размахов сигнализирует о том, что изменение количества звонков оказалось резким, скачкообразным и нехарактерным для данного маршрута. Если выход точки за границы произошел на обеих картах, количество звонков в час изменилось резко и превысило ожидаемые значения, на маршруте произошел всплеск трафика. В свою очередь, это говорит о нарушении процесса – появлении мошеннического трафика на маршруте или других аномалий в трафике. Кроме выхода точки за контрольные границы, в качестве критерия нарушения процесса можно рассматривать формирование на карте серии точек специального вида: группы из семи последовательно возрастающих или убывающих точек подряд, группы из 14 точек в шахматном прядке, резких скачков и т.п. [3-5].

На рис. 4 показаны контрольные карты индивидуальных значений и скользящих размахов для временного ряда количества звонков по рис. 3, который был преобразован по предлагаемому алгоритму.

ВЫВОДЫ

Предложенный алгоритм обнаружения нарушений процесса передачи голосового трафика был проверен на 360 срезах трафика, из которых 77 срезов имели нарушения процесса передачи в виде повышения количества звонков в час. Сравнение производилось с изначальным алгоритмом обнаружения нарушений, не включавший в себя этапы 1 и 3, результаты работы алгоритмов представлены в таблице 1.

Процент ошибки снизился на 2,78% – с 9,17% до 6,39%. Хотя количество пропущенных нарушений процесса на уровне модуля обнаружения аномального трафика увеличилось в 3 раза, количество ложных обнаружений уменьшилось в 5,4 раза. Так как в антифрод-системе для обнаружения мошенничества используются не только контрольные карты, но и статистические характеристики срезов трафика, это может компенсировать количество пропущенных срезов с мошенническим трафиком на уровне антифрод-системы. Хотя полнота обнаружения снизилась на 15,58%, точность обнаружения увеличилась на 19,74%, а F-мера предложенного алгоритма выше, чем у изначального алгоритма на 0,0255 (0,8369 против 0,8114). Предложенный алгоритм возможно дорабатывать, применяя более совершенные методы сегментации и критерии для сравнения выборок.

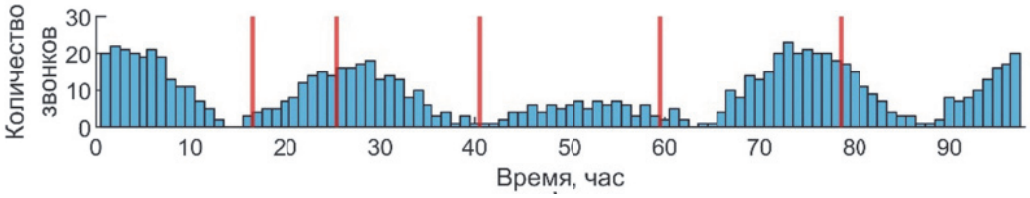

Рисунок 3. Временной ряд количества звонков

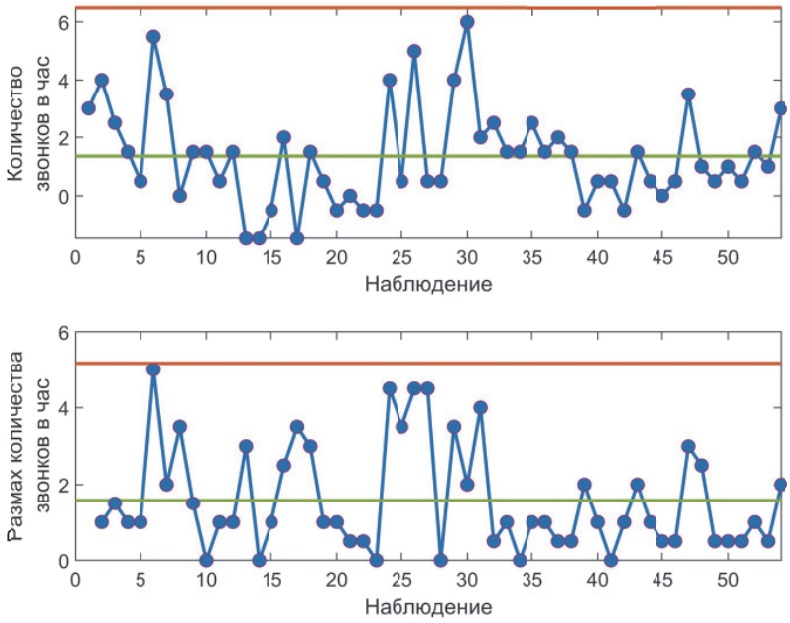

Рисунок 4. Карты индивидуальных наблюдений и скользящих размахов

Таблица 1. Результаты работы изначального (слева) и предложенного (справа) алгоритмов

	Нарушение (контрольные карты)	Норма (контрольные карты)
Нарушение (аналитик)	71	6
Норма (аналитик)	27	256

	Нарушение (контрольные карты)	Норма (контрольные карты)
Нарушение (аналитик)	59	18
Норма (аналитик)	5	278

Кроме мошеннических атак аномалии в трафике вызывает, например, изменение пропускной способности маршрута со стороны других операторов. В этом случае анализ контрольными картами при помощи вышеописанного алгоритма также будет давать ложные срабатывания – ошибки первого рода при обнаружении мошенничества. Поэтому все CDR из часового среза с выявленным всплеском передаются в следующий модуль системы – модуль анализа, который проверяет принадлежность каждого звонка к мошенническому трафику. Такой подход позволяет снизить количество ошибок первого рода за счет последовательной обработки данных вначале статистическими методами, а потом методами машинного обучения. После модуля анализа мошеннические телефонные номера отправляются в модуль блокирования, который выделяет диапазоны телефонных номеров, на которые не будет пропускаться телефонный трафик [13].

СПИСОК ЛИТЕРАТУРЫ

1. Communications Fraud Control Association (CFCA). Telecommunications fraud increased 12% in 2023, equating to an estimated $38.95 billion lost to fraud [Электронный ресурс]. – URL: https://web.archive.org/web/20220314060801/https://cfca.org/wp-content/uploads/2021/02/CFCA-2019-Fraud-Loss-Survey.pdf, 2019 (дата обращения: 28.08.2024).
2. *Лукин, В.Н.* Оценка стабильности циклических процессов с использованием кон-трольных карт Шухарта / В.Н. Лукин, В.В. Ященко // Известия СПбГЭТУ «ЛЭТИ». – 2013. – № 5.
3. *Клячкин, В.Н.* Модели и методы статистического контроля многопараметрического технологического процесса / В.Н. Клячкин. – М.: ФИЗМАТЛИТ, 2011. – 196 с.
4. *Клячкин, В.Н.* Прогнозирование и диагностика стабильности функционирования технических объектов / В.Н. Клячкин, В.Р. Крашенинников, Ю.Е. Кувайскова. – М.: РУСАЙНС, 2020. – 200 с.
5. *Уилер, Д.* Статистическое управление процессами: Оптимизация бизнеса с использованием контрольных карт Шухарта [пер. с англ]. – 3 е изд. / Д. Уилер, Д. Чамберс – М.: Альпина Паблишер, 2023. – 409 с.
6. *Truong, Ch., Oudre, L., Vayatis, N.* A review of change point detection methods. // arXiv. – 2018. – DOI:10.48550/arXiv.1801.00718.
7. *Ermshaus, A., Schäfer, P., Leser, U.* ClaSP: parameter-free time series segmentation // Data Mining and Knowledge Discovery. – 2023. – Vol. 37. – P. 1262–1300. – DOI:10.1007/s10618-023-00923-x.
8. *Ermshaus A., Schäfer P., Leser U.* Raising the ClaSS of Streaming Time Series Segmentation // arXiv. – 2023. – DOI:10.48550/arXiv.2310.20431.
9. ESS Guidelines on Seasonal Adjustment [Электронный ресурс]. – Режим доступа: https://ec.europa.

eu/eurostat/documents/ 3859598/6830795/KS-GQ-15-001-EN-N.pdf, 2015 (дата обращения: 28.08.2024).
10. *Губанов, В.А.* Сравнение методов сезонной корректировки временных рядов / В.А. Губанов // Научные труды: Институт народнохозяйственного прогнозирования РАН. – 2010. – № 8.
11. *Айвазян, С.А.* Прикладная статистика и основы эконометрики / С.А. Айвазян, В.С. Мхитарян. – М.: ЮНИТИ, 1998. – 1022 с.
12. *Karagöz, D., Canan, H.* Control charts for skewed distributions: Weibull, Gamma, and Lognormal // Metodoloski Zvezki. – 2012. – Vol. 9.
13. *Санталов, А.А.* Алгоритм автоматической блокировки диапазонов мошеннических и спамовых телефонных номеров / А.А. Санталов // Вестник Ульяновского государственного технического университета. – 2023. – № 1(101).

USAGE OF CONTROL CHARTS TO FRAUD DETECTION IN THE VOICE TRAFFIC TRANSIT

© 2024 A.A. Santalov[1], V.N. Klyachkin[2]

[1] Telesystems LLC, Saint-Petersburg, Russia
2 Ulyanovsk State Technical University, Ulyanovsk, Russia

The continuous evolution of fraudulent telecommunication traffic necessitates the improvement of anti-fraud systems designed to prevent spam and fraud. The anomaly detection module in such systems utilizes control charts for monitoring, which are plotted based on time series of voice traffic transmission characteristics. These time series are non-stationary, cyclic, and asymmetrically distributed, complicating their analysis. This paper presents an algorithm that reduces the frequency of false positives in control charts and enhances the overall efficiency of anti-fraud systems. The algorithm involves preparing the original time series of voice traffic transmission characteristics for control chart analysis by segmenting the non-stationary series with a cyclic component and subsequently merging homogeneous segments into a single sample.

Keywords: control charts for individual observations and moving ranges, time series segmentation, non-stationary time series, anti-fraud system, monitoring.

DOI: 10.37313/1990-5378-2024-26-4(3)-395-399
EDN: MXBPTB

REFERENCES

1. Communications Fraud Control Association (CFCA). Telecommunications fraud increased 12% in 2023, equating to an estimated $38.95 billion lost to fraud [Elektronnyj resurs]. – URL: https://web.archive.org/web/20220314060801/https://cfca.org/wp-content/uploads/2021/02/CFCA-2019-Fraud-Loss-Survey.pdf, 2019 (data obrashcheniya: 28.08.2024).
2. *Lukin, V.N.* Ocenka stabil'nosti ciklicheskih processov s ispol'zovaniem kon-trol'nyh kart Shuharta / V.N. Lukin, V.V. YAshchenko // Izvestiya SPbGETU «LETI». – 2013. – № 5.
3. *Klyachkin, V.N.* Modeli i metody statisticheskogo kontrolya mnogoparametricheskogo tekhnologicheskogo processa / V.N. Klyachkin. – M.: FIZMATLIT, 2011. – 196 s.
4. *Klyachkin, V.N.* Prognozirovanie i diagnostika stabil'nosti funkcionirovaniya tekhnicheskih ob»ektov / V.N. Klyachkin, V.R. Krasheninnikov, YU.E. Kuvajskova. – M.: RUSAJNS, 2020. – 200 s.
5. *Uiler, D.* Statisticheskoe upravlenie processami: Optimizaciya biznesa s ispol'zovaniem kontrol'nyh kart Shuharta [per. s angl]. – 3 e izd. / D. Uiler, D. Chambers – M.: Al'pina Pablisher, 2023. – 409 s.
6. *Truong, Ch., Oudre, L., Vayatis, N.* A review of change point detection methods. // arXiv. – 2018. – DOI:10.48550/arXiv.1801.00718.
7. *Ermshaus, A., Schäfer, P., Leser, U.* ClaSP: parameter-free time series segmentation // Data Mining and Knowledge Discovery. – 2023. – Vol. 37. – P. 1262–1300. – DOI:10.1007/s10618-023-00923-x.
8. Ermshaus A., Schäfer P., Leser U. Raising the ClaSS of Streaming Time Series Segmentation // arXiv. – 2023. – DOI:10.48550/arXiv.2310.20431.
9. ESS Guidelines on Seasonal Adjustment [Elektronnyj resurs]. – Rezhim dostupa: https://ec.europa.eu/eurostat/documents/ 3859598/6830795/KS-GQ-15-001-EN-N.pdf, 2015 (data obrashcheniya: 28.08.2024).
10. *Gubanov, V.A.* Sravnenie metodov sezonnoj korrektirovki vremennyh ryadov / V.A. Gubanov // Nauchnye trudy: Institut narodnohozyajstvennogo prognozirovaniya RAN. – 2010. – № 8.
11. *Ajvazyan, S.A.* Prikladnaya statistika i osnovy ekonometriki / S.A. Ajvazyan, V.S. Mhitaryan. – M.: YUNITI, 1998. – 1022 s.
12. *Karagöz, D., Canan, H.* Control charts for skewed distributions: Weibull, Gamma, and Lognormal // Metodoloski Zvezki. – 2012. – Vol. 9.
13. *Santalov, A.A.* Algoritm avtomaticheskoj blokirovki diapazonov moshennicheskih i spamovyh telefonnyh nomerov / A.A. Santalov // Vestnik Ul'yanovskogo gosudarstvennogo tekhnicheskogo universiteta. – 2023. – № 1(101).

Anton Santalov, Systems Analyst of the Statistical Research and Development Department, Postgraduate Student of the Department of Applied Mathematics and Informatics of the Ulyanovsk State Technical University (UlSTU).

E-mail: anton.santalov1995@gmail.com
Vladimir Klyachkin, Doctor of Technical Sciences, Professor of the Department of Applied Mathematics and Informatics of UlSTU. E-mail: v_kl@mail.ru