УДК 004.89 DOI: 10.26102/2310-6018/2020.29.2.011

Интеллектуальный анализ данных пользовательского окружения в задаче обнаружения удаленного управления

А.М. Вульфин
ФГБОУ ВО «Уфимский государственный авиационный технический университет»
Уфа, Российская Федерация

Резюме: Целью работы является совершенствование алгоритмов обнаружения удаленного управления пользовательским сеансом. Объект исследования – система обнаружения удаленного управления компьютером пользователя. Предмет исследования – алгоритмы интеллектуального анализа данных, собираемых с помощью инструментов и средств мониторинга в составе клиентской части Web-приложения на стороне браузера, предназначенные для анализа изменения паттернов динамических биометрических признаков в случае удаленного управления. Проанализированы подходы к обнаружению удаленного подключения. Разработана структура системы обнаружения удаленного доступа с современным подходом к сбору и анализу пользовательского окружения в сочетании с методами машинного обучения. Экспериментальная часть работы выполнена на основе анализа базы данных пользовательского окружения, собранных специально для тестирования программной реализации разработанных алгоритмов. Было рассмотрено 16 различных вариантов удаленного подключения с злоумышленника к устройству пользователя. Полученная выборка включала 178 замеров с разным количеством временных интервалов между промежуточными точками траектории движения курсора мыши. Наибольшую эффективность показал алгоритм классификации «случайный лес» с группой признаков, состоящих из временных интервалов между событиями движения курсора мыши. Доля верных предсказаний составила 93 % на тестовых данных.

Ключевые слова: интеллектуальный анализ, анализ пользовательского окружения, антифрод система, кибермошенничество, удаленный доступ

Для цитирования: Вульфин А.М. Интеллектуальный анализ данных пользовательского окружения в задаче обнаружения удаленного управления. *Моделирование, оптимизация и информационные технологии.* 2020;8(2). Доступно по: https://moit.vivt.ru/wp-content/uploads/2020/05/Vulfin_2_20_2.pdf DOI: 10.26102/2310-6018/2020.29.2.011

Data mining the user's environment in the problem of remote control detection

A.M. Vulfin
Ufa State Aviation Technical University, Ufa, Russian Federation

Abstract: The aim of the work is to improve the detection algorithms for remote control of a user session. Object of study - a system for detecting remote control of a user's computer. The subject of the study is data mining algorithms collected using tools and monitoring tools as part of the client side of the web application on the browser side, designed to analyze changes in the patterns of dynamic biometric features in the case of remote control. The approaches to detecting a remote connection are analyzed. The structure of the remote access detection system with a modern approach to the collection and analysis of the user environment in combination with machine learning methods has been developed. The experimental part of the work is based on an analysis of the user environment database, collected specifically for testing the software implementation of the developed algorithms. 16 different options for remote connection from an attacker to a user's device were considered. The obtained sample included 178 measurements with a different number of time intervals between intermediate points of the mouse cursor path. The highest efficiency was shown by the random forest classification algorithm with a group

of features consisting of time intervals between mouse cursor movement events. The share of correct predictions was 93% on test data.

Keywords: intelligent analysis, user environment analysis, antifraud system, cyber fraud, remote access

For citation: Vulfin A.M. Data mining the user's environment in the problem of remote control detection. *Modeling, optimization and information technology*. 2020;8(2). https://moit.vivt.ru/wp-content/uploads/2020/05/Vulfin_2_20_2.pdf DOI: 10.26102/2310-6018/2020.29.2. 011 (In Russ).

Введение

Следствием стремительного развития финансовых технологий во всем мире стало возросшее количество киберпреступлений и мошеннических действий в информационной сфере. Важным аспектом функционирования цифровой экономики является обеспечение кибербезопасности бизнеса и защиты персональных данных пользователей.

По данным МИД РФ [1] ущерб мировой экономике от киберпреступности в 2019 г. может достичь $2 трлн., а в 2020 году – до $3 трлн. За 2018 г. по данным МВД РФ число преступлений с использованием современных информационно-телекоммуникационных технологий в РФ увеличилось на 92 %. По данным Центрального банка РФ доля мошеннических операций в интернет-банкинге за последние 2 года выросла в 5,5 раз и составила 93 % всех преступлений, связанных с хищением средств со счетов держателей карт [2]. Число интернет-краж с банковских карт выросло до 300 тыс., при этом размер суммарного ущерба в 2018 г. составил 1,4 млрд руб. согласно статистике ЦБ РФ.

В качестве основных способов реализации киберпреступлений в сфере защиты персональных данных Роскомнадзор выделяет использование вредоносного программного обеспечения на устройстве пользователя и использование методов социальной инженерии. Главной задачей этих мошеннических технологий является получение персональных идентификационных или аутентификационных данных клиента в системах дистанционного банковского обслуживания (ДБО).

Анализ Интернет-ресурсов по информационной безопасности и активное обсуждение проблем безопасности информации в сети Интернет показал, что одной из актуальных существующих мошеннических схем в данной области является удаленное подключение злоумышленника к компьютеру пользователя и дальнейшее управление им в незаконных целях [3].

На сегодняшний день основные сервисы электронной коммерции перешли в Web приложения, что значительно усложняет обнаружение удаленного подключения к компьютеру пользователя из-за высокого уровня абстракции запускаемого в браузере кода (многоуровневая виртуализация). Таким образом, актуальным является разработка алгоритмов обнаружения удаленного управления пользовательским сеансом с помощью инструментов и средств мониторинга в составе клиентской части Web-приложения на стороне браузера. Перспективным является анализ изменения паттернов динамических биометрических признаков в случае удаленного управления сеансом.

Цель работы: совершенствование алгоритмов обнаружения удаленного управления пользовательским сеансом.

1. Анализ существующих подходов к обнаружению удаленного управления

Для создания удаленного подключения используют специализированное программное обеспечение, позволяющее выполнять все действия на удаленном устройстве, изменять настройки ПО, обмениваться файлами, делать снимки экрана,

шифровать передаваемые данные, проводить конференции, подключать Web-камеры, удаленные проекторы и прочие сетевые устройства [4].

На сегодняшний день самыми популярными компьютерными операционными системами являются Windows и Linux, а для смартфонов – iOS и Android. Поддержка операционных систем и особенности систем удаленного доступа представлены в Таблице 1.

Таблица 1 – Обзор программных систем организации удаленного доступа
Table 1 – Overview of software systems for organizing remote access.

Название программного средства	Поддерживаемые ОС				Особенности
	Windows	Linux	Android	IOS	
Teamviewer	+	+	+	+	простота работы, запуска и настройкой
Ammy admin	+	-	-	-	
Radmin	+	-	-	-	состоит из: клиентской (Radmin Viewer) и серверной (Radmin Server) части.
Anydesc	+	+	+	+	
Trustviewer	+	-	-	-	
Supremo	+	-	+	+	

1.1 Анализ способов обнаружения удаленного управления

Способы, позволяющие обнаружить сеанс удаленного управления:

- сканирование открытых портов позволяет обнаруживать только те средства удаленного управления, которые используют фиксированные порты. Например, Radmin, имеет стандартный порт по умолчанию (4899), предоставляет возможность настройки порта в диапазоне от 1 до 65535. Таким образом, если используется один из широко известных портов, то обнаружить Radmin будет сложнее. TeamViewer использует 80 и 443 порты HTTP. Ammyy Admin использует HTTPs прокси, поэтому обнаружить его сканированием открытых портов невозможно;

- сканирование реестра. Можно обнаружить, например, TeamViewer, Ammyy Admin и Radmin.

- сканирование сетевого трафика. Многие средства удаленного управления шифруют свой трафик. Radmin использует известные протоколы удаленного управления без шифрования. TeamViewer и Ammyy Admin шифруют трафик.

- сигнатурный анализ.

Самым доступным для пользователя средством защиты от программ удаленного управления является антивирус. Известные антивирусы, в частности Kaspersky Internet Security [5], обнаруживают исполняемые файлы, сканируют оперативную память, службы, записи в реестре, характерные для средств удаленного управления, и не позволяют работать подобным средствам без разрешения пользователя. Dr.Web обнаруживает средства удаленного управления не только сигнатурным анализом, но и при проверке запущенных служб. Например, Ammyy Admin блокируется многими

известными антивирусами. Однако, если внести подобные средства в исключения, то их работа не будет блокироваться.

Несмотря на то, что антивирусы, при их относительной сложности обнаруживают наличие средств удаленного управления, сам факт установления сеанса удаленного управления они не обнаруживают [3].

Более того, если имеется возможность использовать для анализа только данные, получаемые из Web окружения приложения, то задача значительно усложняется [6-16].

1.2 Существующие системы мониторинга транзакций

Система мониторинга транзакций (СМТ) или антифрод система – специализированный программный или программно-аппаратный комплекс, обеспечивающий мониторинг, обнаружение мошеннических действий, а также обеспечивающий поддержку принятия решения по обнаруженной незаконной операции [17]. Антифрод системы, которые способны детектировать средства удаленного управления, представлены в Таблице 2.

Таблица 2 – Обзор СМТ, детектирующих средства удаленного управления
Table 2 – Overview of systems that detect remote controls.

Наименование системы	Поиск признаков дискредитации и во внешних источниках	Использование методов машинного обучения	Анализ окружения пользователя
Kaspersky Fraud Prevention, Kaspersky, Россия	-	+	+
RSA Transaction Monitoring, Инфосистемы Джет, Россия	-	+	-
ICFraud, Frodex, Россия	+	+	+
Bot-Trek Intelligent Bank, 4by4, Россия	+	-	-

СМТ на основе методов интеллектуального анализа данных хорошо зарекомендовали себя в сфере обработки экономических и банковских данных: различные реализации нейронных сетей [18, 19], метода опорных векторов (Support Vector Machine – SVM) [20], скрытые Марковские модели (Hidden Markov Model – HMM) [21-23], генетические алгоритмы и эволюционное программирование [18, 24, 25], деревья решений (Decision Trees) [26], нечеткая логика (Fuzzy Logic) [19]. Однако в базовом варианте эти алгоритмы уже не способны решить существующие задачи в виду возрастающих объемов обрабатываемых данных. Возникает потребность модифицировать эти алгоритмы, а также комбинировать их для получения приемлемого результата. Технологии построения ансамблей классификаторов (бэггинг (bagging)), комитетов слабых классификаторов (бустинг (boosting) [23]) и композиции гетерогенных классификаторов (стэкинг (stacking) [24]) позволяют достичь высоких показателей в задаче анализа банковских данных. Кроме того, в сфере интернет-банкинга актуальным становится применение технологий больших данных (Big Data) [27, 28], позволяющие обрабатывать огромные объемы накапливаемой информации. Для повышения эффективности антифрод систем многие разработчики не ограничиваются собственными черными листами мошенников. В таких системах получение информации

о том или ином клиенте возможно из сторонних доверенных источников, что снижает вероятность пропуска мошеннических операций.

Достаточно новым способом идентификации пользователя является анализ пользовательского окружения [6, 11-16] – собирается информация не только об IP-адресе устройства клиента, но и различные характеристики, идентифицирующие устройство клиента: от размера от экрана устройства до списка установленных шрифтов. Такой способ идентификации стал актуальным из-за учащающихся случаев совершения мошеннических действий путем удаленного управления компьютером клиента, а также незаконного получения идентификационных данных клиента (например, логина или пароля), то есть неизвестно, клиент совершил операцию или нет.

1.3 Обзор алгоритмов анализа динамических биометрических признаков как инструмента выявления удаленного управления

Попытки обнаружить и классифицировать динамические биометрические шаблоны предпринимались рядом отраслевых игроков. Система обнаружений удаленного подключения выполняет сложные действия по извлечению признаков, измерению и нормализации данных, характеризующих работу пользователя за ПК – движения курсора мыши, клавиатурный почерк. Для каждого события передвижения курсора мыши система использует искусственное сглаживание траектории, фиксирует траекторию движения, измеряя в разных точках скорость, ускорение, кривизну, относительные расстояния, точки изменения траектории движения и т. д. После такой предварительной обработки к извлеченным признакам применяются методы машинного обучения. Модель обучается, затем осуществляется прогнозирование.

Для решения задачи обнаружения удаленного управления в [7] данные о движении курсора мыши были объединены и преобразованы в цветные изображения. Одно такое изображение представляет траекторию и координаты тысячи точек активности курсора мыши. Был разработан специальный высококонтрастный алгоритм цветового кодирования для согласованного представления направления, скорости и ускорения движений курсора.

В [9] рассмотрена проблема идентификации оператора по виброакустическому сигналу, возникающему при работе с компьютерной мышью

В [8] описано выявления удаленного подключения на основе данных о событиях и прерываниях компьютерной мыши. Способ при работе на страницах Web-ресурса включает этапы, на которых предварительно собирают данные о периодичности срабатывания события движения компьютерной мыши, получают события движения компьютерной мыши, сравнивают периодичность срабатывания полученного события движения компьютерной мыши с вышеуказанной сформированной статистической моделью, при возникновении отклонения в срабатывании события движения компьютерной мыши уведомляют владельца защищаемого Web-ресурса о наличии удаленного подключения у посетителя Web-ресурса для последующего реагирования на стороне владельца.

2. Разработка структуры системы обнаружения удаленного подключения в среде виртуального окружения Web-браузера

Технологии дистанционного банковского обслуживания (ДБО) для доступа к счетам и операциям через Web-браузер не требуют установки клиентской части ПО и получили очень широкое распространение [29, 30]. Стандартный сценарий работы с системой ДБО включает следующие этапы. Пользователь совершает определенные манипуляции в Web-браузере, взаимодействующим с интерфейсом Frontend-сервера. Frontend-сервер передает данные о действиях пользователя по инициализации

транзакции на Backend-сервер системы дистанционного банковского обслуживания (ДБО) и затем в автоматизированную банковскую систему (АБС) банка для проведения расчетов [29, 30]. Backend сервер передает данные транзакции для анализа в антифрод-систему. В случае признания легитимности транзакции данные передаются на сервера АБС, в противном случае Backend-сервер отказывает пользователю в совершении операции. В случае применения технологий получения пользовательского окружения, описанных выше, на Frontend-сервера банка помещают скрипт. На этапе формирования пользователем платежного поручения внедренный скрипт опрашивает браузер и считывает всю доступную ему информацию. Затем эти данные вместе с данными о транзакции передаются на Backend-сервер для анализа.

Предлагается следующая обобщенная структурная схема системы обнаружения удаленного подключения в среде виртуального окружения Web-браузера на основе анализа динамических биометрических признаков, представленная на Рисунке 1. Введены следующие обозначения:

1) настоящий пользователь;
2) пользователь удаленного подключения;
3) параметры, фиксируемые в браузере на стороне клиента и пересылаемые на сервер (траектория движения мыши + клавиатурный почерк);
4) параметры пользовательской сессии (включая динамические биометрические признаки);
5) параметры, передаваемые в существующую систему анализа параметров пользовательского окружения;

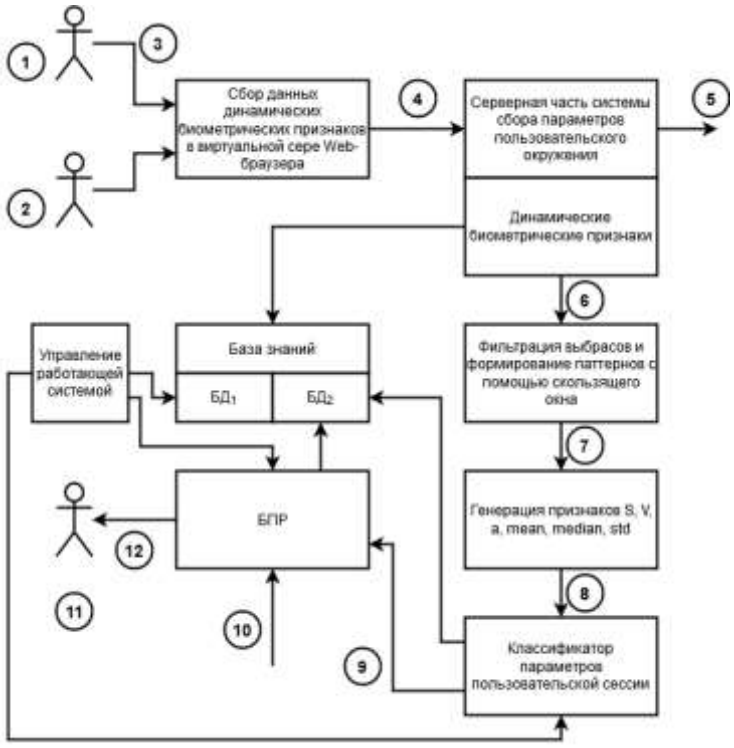

Рисунок 1 – Обобщенная структурная схема системы обнаружения удаленного подключения в среде виртуального окружения Web-браузера на основе анализа динамических биометрических признаков

Figure 1 – The generalized block diagram of a remote connection detection system in a virtual environment of a Web browser based on an analysis of dynamic biometric features.

6) фиксируемые динамические биометрические признаки;
7) подготовленные паттерны динамических биометрических признаков с привязкой к пользовательской сессии;
8) вектор признаков пользовательской сессии;
9) оценка вероятности наличия удаленного подключения;
10) оценка вероятности мошеннических действий из системы в рамках текущей сессии;
11) аналитик антифрод-системы;
12) оценка вероятности мошеннических действий в пределах сессии;

$БД_1$ – БД параметров пользовательской сессии;
$БД_2$ – параметры текущей работы предобработка и классификация.

2.1 Разработка алгоритмов обнаружения удаленного подключения

Для анализа будем использовать динамические биометрические признаки – параметры использования компьютерной мыши, а именно:

- координаты точек промежуточной остановки указателя мыши;
- временные интервалы между началом движения мыши и ее остановкой.

При детальном анализе получившихся временных рядов, характеризующих движения курсора мыши в клиентской части Web-приложения на стороне пользователя (запускаемого в браузере, фиксирующем параметры движения с помощью встроенных средств) было выявлено, что начало каждой такой последовательности имело множество аномальных значений, это связано с длительной загрузкой составных частей страницы в браузере, поэтому было решено отсечь первые 25% значений.

Но критические значения выбросов присутствовали и далее на протяжении всей сессии, сложно сказать точно, с чем это было связано, потому что на это влияет огромное количество факторов.

Далее ряд временных интервалов был представлен в виде гистограммы, характеризующей длительности интервалов и количество таких интервалов. Из гистограммы были выделены те участки траектории движения мыши, временные интервалы которых лежали в диапазоне от 25% до 75% образцов (см. Рисунок 2).

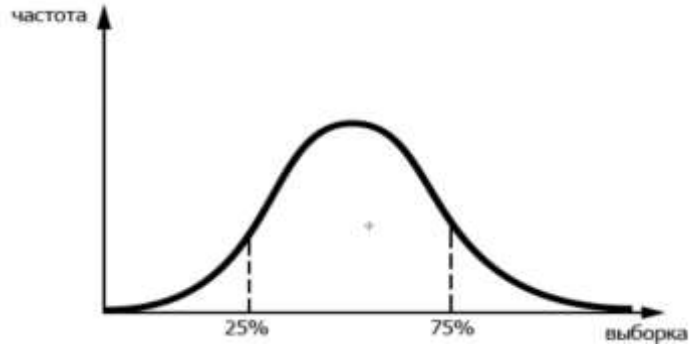

Рисунок 2 – Гистограмма разложения ряда длительности интервалов между событиями.
Figure 2 – The histogram of the expansion of intervals between events durations series.

Результат после данных действий можно увидеть на диаграмме одной случайно выбранной последовательности (см. Рисунок 3).

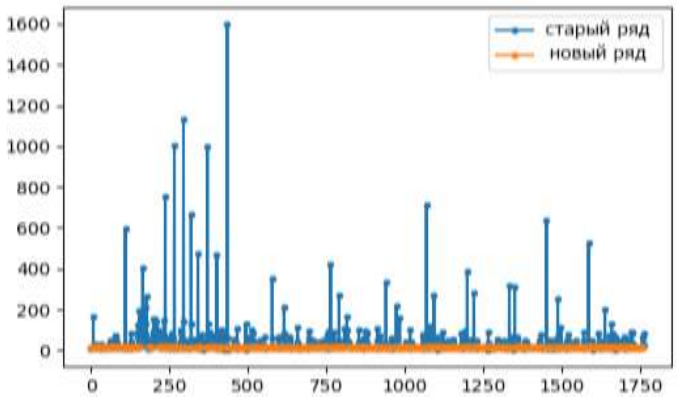

Рисунок 3 – Сравнительная диаграмма исходного и сокращённого ряда временных интервалов
Figure 3 – Comparative chart of the initial and shortened series of time intervals.

Так как количество точек (X, Y, временной интервал) в каждом ряду разное, что обусловлено уникальными характеристиками перемещений мыши человеком, то необходимо привести длины рядов к одной величине, чтобы их можно было использовать в качестве вектора признаков для подачи в классификатор. Для этого используем скользящее окно с фиксированным шагом по нормализованному набору интервалов длительностей, это позволит увеличить размер самой выборки и получить множество паттернов.

Для имитации траектории движения пользователя построим кубический сплайн по промежуточным точкам остановки курсора мыши, фиксируемых внутри страницы Web-браузера. Применение кусочно-линейной интерполяции не позволяет приблизится к реальным траекториям движения курсора мыши и качественно оценивать длины соответствующих участков траектории.

Становится возможным определить скорости по формуле (1) и ускорения по формуле (2) на каждом временном интервале между двумя соседними точками траектории движения мыши.

$$v_i = \frac{l_i}{t_i}, \tag{1}$$

где l_i – расстояние между i и i–1 точками сглаженной траектории;

t_i – временной отсчет в точке i.

$$a_i = \frac{dv_i}{dt_i}, \tag{2}$$

где v_i – скорость в i-ой точке сплайна;

t_i – временной отсчет в точке i.

Найдем среднее значение для временных интервалов по выделенному паттерну по формуле (3):

$$\overline{t} = \frac{1}{n}\sum_{i=1}^{n} t_i, \tag{3}$$

где n – количество точек в ряду;

t_i – временной отсчет в точке i.

По формуле (3) найдем также средние значения для S, v, a.

Найдем медианы для временных интервалов. Все значения отсортируем по убыванию. Медиану для ряда, состоящего из четного количества элементов найдем по формуле (4):

$$Me_t = \frac{t_{\frac{N}{2}} + t_{\frac{N+1}{2}}}{2}, \quad (4)$$

где N – количество значений в ряду.

Медиану для ряда, состоящего из нечетного количества элементов найдем по формуле (5):

$$Me_t = x_{\frac{N+1}{2}}, \quad (5)$$

где N – количество значений в ряду.

Найдем стандартные отклонения для временных интервалов по формуле (6):

$$std_t = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(t_i - \bar{t}_i\right)^2}, \quad (6)$$

где n – количество точек в ряду.

Таким образом можно построить иерархию признаков, характеризующих участок траектории движения курсора мыши в окне Web-приложения, которые фиксируются и передаются в CMT для последующего анализа. Финальный вектор признаков представлен в Таблице 3 и на Рисунке 4.

Таблица 3 – Финальный вектор признаков
Table 3 – Final Feature Vector

Временные отсчеты промежуточных событий движения курсора мыши	t	длительность одного сегмента траектории движения
	mean(t), $\bar{t}$	среднее значение длительности одного сегмента
	std(t), std_t	отклонение
	median(t), Me_t	медиана
Координаты точек промежуточных событий движения курсора мыши	s	длина одного сегмента траектории движения
	mean(s), $\bar{s}$	среднее значение длины одного сегмента
	std(s), std_s	отклонение
	median(s), Me_s	медиана
Относительная скорость движения курсора мыши между двумя промежуточными событиями движения	v	относительная скорость движения курсора мыши в сегменте траектории движения
	mean(v), $\bar{v}$	среднее
	std(v), std_v	отклонение
	median(v), Me_v	медиана
Ускорения движения курсора мыши между двумя промежуточными событиями движения	a	ускорение курсора мыши в сегменте траектории движения
	mean(a), $\bar{a}$	среднее
	std(a), std_a	отклонение
	median(a), Me_a	медиана

t	S	v	a	$\bar{t}$	$\bar{s}$	$\bar{v}$	$\bar{a}$	Me_t	Me_s	Me_v	Me_a	std_t	std_s	std_v	std_a

Рисунок 4 – Финальный вектор признаков
Figure 4 – Final Feature Vector

В качестве классификаторов будут использованы:

– KNN (классификатор *k*-ближайших соседей);

– RF (классификатор на основе комитета случайных деревьев, Random Forest).

Имеется два класса («удаленное управление» и «работа настоящего пользователя») и алгоритм, определяющий принадлежность каждого объекта одному из классов, тогда матрица ошибок классификации представлена в Таблице 4.

Таблица 4 – Матрица ошибок
Table 4 – Error matrix

	$y = 1$	$y = 0$
$y = 1$	True Positive (TP)	False Positive (FP)
$y = 0$	False Negative (FN)	True Negative (TN)

Здесь y – это ответ алгоритма на объекте, а y – истинная метка класса на этом объекте. Для оценки результатов классификации, как правило, применяются метрики: Accuracy, Precision, Recall и F_1-мера, которые основываются на основных показателях классификации [31]:

– истинно положительные (TP). Мошеннические операции, классифицированные, как мошеннические;

– истинно отрицательные (TN). Добросовестные операции, классифицированные, как добросовестные;

– ложно положительные (FP). Добросовестные операции, классифицированные, как мошеннические;

– ложно отрицательные (FN). Мошеннические операции, классифицированные, как добросовестные.

Таким образом, ошибки классификации бывают двух видов: Ложно отрицательные (FN) и Ложно Положительные (FP).

Для оценки эффективности были использованы следующие метрики:

– Accuracy – доля правильных ответов алгоритма. Находится по формуле (7):

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

– Precision (отражает долю объектов, названных классификатором положительными и при этом действительно являющимися положительными). Находится по формуле (8):

$$precision = \frac{TP}{TP + FP} \tag{8}$$

– Recall (показывает, какую долю объектов положительного класса из всех объектов положительного класса нашел алгоритм).Находится по формуле (9):

$$recall = \frac{TP}{TP + FN} \tag{9}$$

– F_1 – среднее гармоническое precision и recall. Находится по формуле (10):

$$F_\beta = \left(1 + \beta^2\right) \cdot \frac{precision \cdot recall}{\left(\beta^2 \cdot precision\right) + recall}, \tag{10}$$

где β – вес точности, равный 1.

Основными этапами в задаче машинного обучения являются следующие этапы:

1) сбор данных;
2) анализ и обработка данных;
3) выбор модели и ее обучение;
4) оценка качества построенной модели.

Сбор данных о пользовательском окружении осуществляется посредством скрипта JavaScript. Далее следует этап анализа и обработки данных. На этом этапе решаются задачи:

– features selection – отбор наиболее значимых признаков для принятия классификационного решения;
– features transformation – заполнение пропусков в данных, удаление выбросов и фильтрация шумовых компонент;
– features extraction – преобразование отобранных признаков в новое пространство признаков для подачи на вход классификатора.

3. Оценка эффективности алгоритма выявления удаленного управления и анализ результатов

3.1 Разработка стенда для сбора данных

Для того чтобы разработать методику сбора данных для алгоритма анализа данных пользовательского окружения, необходимо практически воссоздать условия, приближенные к реальным.

Схема работы для нашего случая выглядит следующим:

– Устройство 1. Это устройство злоумышленника, оно может быть необязательно стационарным компьютером, но также и смартфоном, и планшетом, и ноутбуком. Будем рассматривать только стационарный компьютер и смартфон, так как данные устройства являются наиболее используемыми и доступными.

– Устройство 2. Это рабочая станция пользователя. В качестве такого устройства будем использовать стационарный компьютер, так как это приближает нас к реальным условиям.

– Сервер. Это имитация посещения сайта банка. При переходе по IP адресу сервера открывается Web-форма, которая позволяет перевести условные деньги и таким образом собрать необходимые данные о действиях пользователя на странице. В качестве операционной системы и базы данных для сервера были выбраны Debian 8.1 и Postgesql, так как они оптимально подходят для данных задач.

Злоумышленник с Устройства 1 подключается к Устройству 2 пользователя, переходит по IP-адресу сервера, попадает на Web-форму банка и заполняет её. Таким образом, мы можем собрать достаточно размеченных данных для последующего анализа.

3.2 Методика сбора данных

В качестве Устройства 1, будем использовать стационарный компьютер и смартфон. Но подключение с ПК может осуществляться с разных операционных систем и нужно понимать, как та или иная ОС может влиять на наши данные, также подключение может осуществляться с виртуальной машины, и в таком случае тоже имеется возможность произвести подключение с различных операционных систем. Поэтому было решено рассматривать 3 типа подключения:

- Из ОС Windows 7;
- Из виртуальной машины на Windows 7;
- Из виртуальной машины на Linux (Debian 8.1).

В качестве операционных систем для Устройства 2 пользователя мы будем использовать Windows 7 и Windows XP.

Удаленное подключение может осуществляться при помощи разного программного обеспечения:

- TeamViewer;
- Radmin;
- Ammyy Admin;
- AnyDesc;
- Supremo Remote Desktop;
- Trust Viewer.

Не все программы разработаны одновременно для операционных систем Windows и Linux. Поэтому набор приложений для каждой ОС будет отличаться.

Для операционной системы Windows 7:

- TeamViewer;
- Radmin;
- Ammyy Admin;
- AnyDesc;
- Supremo Remote Desktop.
- Trust Viewer.

Для операционной системы Linux:

- TeamViewer;
- AnyDesc.

Также будем использовать в качестве устройства злоумышленника смартфон. На данный момент наиболее популярны устройства под управлением операционных систем Android и iOS.

На эти ОС можно установить:

- TeamViewer;
- Supremo Remote Desktop;
- Anydesc.

Таким образом:

- 6 подключений с ПК на Windows;
- 6 подключений с ПК на виртуальной машина с Windows 7;
- 2 подключения с ПК на виртуальной машина с Linux;
- 2 подключения со смартфона на Android/IOS.

Всего получилось 16 различных вариантов удаленного подключения с Устройства 1 злоумышленника к Устройству 2 пользователя.

3.3 Эксперимент на собранных данных

После проведения сбора данных выборка включала 178 замеров с разным количеством временных интервалов между промежуточными точками траектории движения курсора мыши. Для эксперимента использовались только динамические значения (Координаты X, Y и время перемещения мыши между остановками). Далее все замеры были обработаны скользящим окном по 64 временных интервала в каждом с шагом 10. В итоге получилось 10686 записей.

Были взяты два классификатора с оптимальными гиперпараметрами: случайный лес (RF) и k-ближайших соседей (KNN).

В итоге полный набор признаков получился следующим: временные интервалы (64 признака), кубический сплайн по временным интервалам (64 признака), скорости (63 признака), ускорения (62 признака), медианы и средние значения по временным интервалам, кубическому сплайну по временным интервалам, скоростям и ускорениям (8 признаков).

Для обучения были использованы разные группы признаков (см. Таблица 5).

Таблица 5 – Группы признаков
Table 5 – Feature Groups

№	Состав группы		Количество признаков
	Название	Обозначение	
1	Медианы и средние значения для временных интервалов, сплайна, скорости, ускорения	Me_t, Me_s, Me_v, Me_a, $\overline{t}$, $\overline{S}$, $\overline{v}$, $\overline{a}$	8
2	Временные интервалы, сплайн по временным интервалам, скорости, ускорения	*t, S, v, a*	252
3	Временные интервалы	*t*	64
4	Сплайн по временным интервалам	*S*	64
5	Скорости	*v*	63
6	Ускорения	*a*	62

Оценка эффективности

Была использована перекрестная проверка с k=10 заходами.

Видно, что, при использовании классификатора k-ближайших соседей, лучший результат наблюдается у первой группы признаков (см. Рисунок 5).

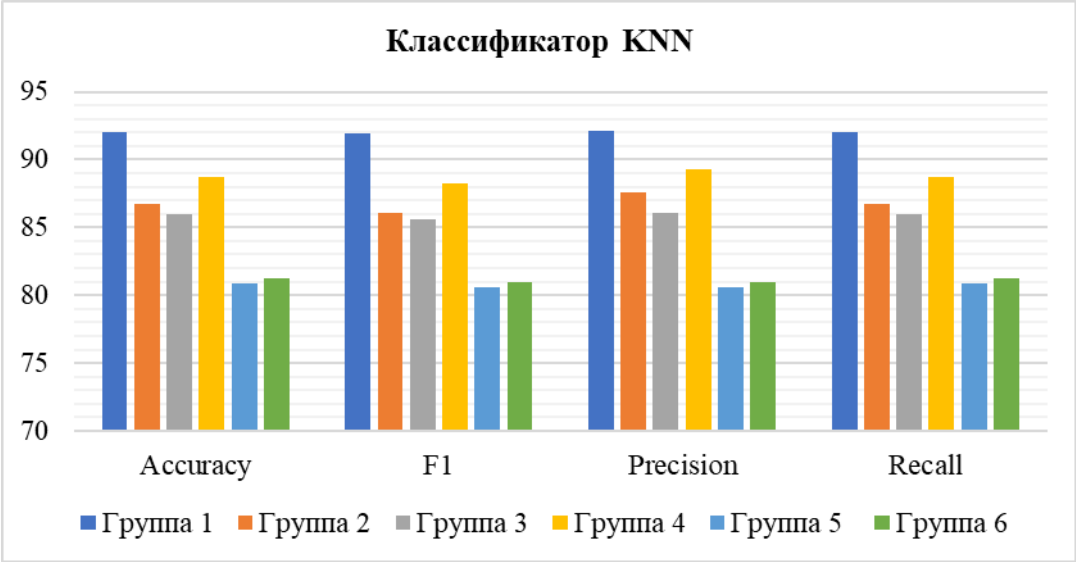

Рисунок 5 – Диаграмма классификатора KNN. По оси ординат показатель качества классификации, %

Figure 5 – KNN classifier diagram. The ordinate is an indicator of the quality of classification, %

У классификатора «случайный лес» лучшим результатом оказывается использование третьей группы признаков (см. Рисунок 6).

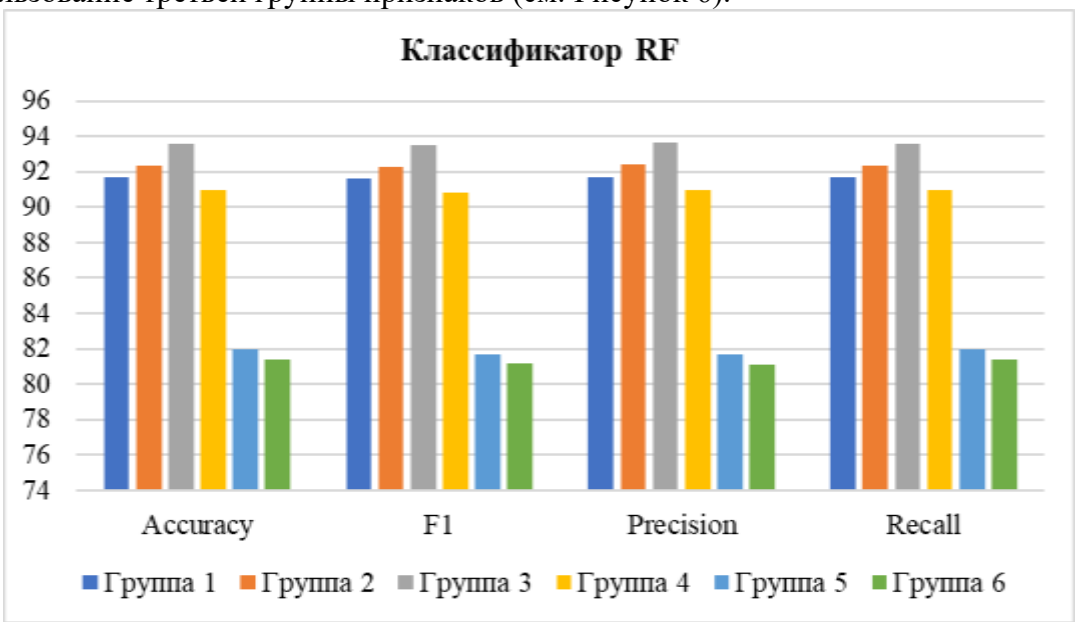

Рисунок 6 – Диаграмма классификатора RF. По оси ординат показатель качества классификации, %

Figure 6 – RF classifier diagram. The ordinate is an indicator of the quality of classification,%

Лучший результат получился с использованием значений только временных интервалов (3 группа – 64 признака) и классификатора Random Forest, точность – 93,61% (см. Рисунок 7).

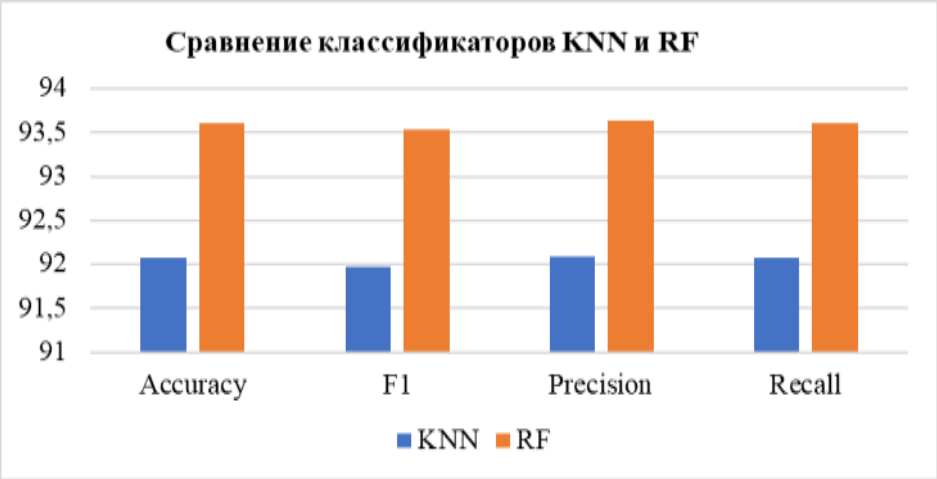

Рисунок 7 – Сравнение классификаторов KNN и RF. По оси ординат показатель качества классификации, %
Figure 7 – Comparison of KNN and RF classifiers. The ordinate is an indicator of the quality of classification, %

Наибольшую эффективность показал алгоритм классификации «случайный лес» с группой признаков, состоящих из временных интервалов. Доля верных предсказаний для этого алгоритма составила 0,9361.

Заключение

Анализ параметров клиентской части Web-приложения на предмет обнаружения удаленного подключения затруднителен из-за высокого уровня абстракции исполняемого в браузере кода. Таким образом, проблемой является разработка алгоритмов обнаружения удаленного управления пользовательским сеансом с помощью инструментов и средств мониторинга в составе клиентской части Web-приложения на стороне браузера. Применен подход на основе анализа изменения паттернов динамических биометрических признаков в случае удаленного управления сеансом.

Выполнен обзор популярных программ удаленного администрирования, проанализированы современные системы детектирования удаленного администрирования, выполнен анализ алгоритмов обработки динамических признаков.

Разработана структура системы обнаружения удаленного доступа с современным подходом к сбору и анализу пользовательского окружения в сочетании с методами машинного обучения. Применение методов машинного обучения позволит автоматизировать процесс адаптации системы к новым схемам мошенничества. Проведен анализ алгоритмов машинного обучения для анализа данных пользовательского окружения, рассмотрены алгоритмы классификации, а также различные метрики качества. Разработана методика сбора данных.

Проведен вычислительный эксперимент на натурных данных. Наибольшую эффективность показал алгоритм классификации «случайный лес» с группой признаков, состоящих из временных интервалов между событиями движения курсора компьютерной мыши. Доля верных предсказаний для этого алгоритма составила 93 % на тестовых данных.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-08-00668

ЛИТЕРАТУРА

1. МИД РФ: ущерб мировой экономике от киберпреступности в 2019 году может достичь $2 трлн [Электронный ресурс]. URL: https://tass.ru/politika/5551244 (дата обращения: 11.04.2020).
2. Мошенничество с банковскими картами [Электронный ресурс]. URL: http://www.tadviser.ru/index.php/Статья:Мошенничество_с_банковскими_картами (дата обращения: 11.04.2020).
3. -Детектирование сеанса удаленного управления методами клавиатурного мониторинга [Электронный ресурс]. URL: http://www.frodex.ru/article/radkl2015 (дата обращения: 11.04.2020).
4. Удаленный доступ к компьютеру и как его организовать: расширяем горизонты бизнеса [Электронный ресурс]. URL: https://www.kp.ru/guide/udalennyi-dostup-k-komp-juteru.html (дата обращения: 11.04.2020).
5. Kaspersky Fraud Prevention: Решение для эффективной защиты от кибермошенничества [Электронный ресурс]. URL: https://www.karma-group.ru/catalog/kaspersky-for-enterprise/fraud-prevention/ (дата обращения: 11.04.2020).
6. Сапожникова М.Ю., Вульфин А.М., Гаянова М.М., Никонов А.В. Алгоритмы интеллектуального анализа данных банковских транзакций в составе системы противодействия финансовому мошенничеству. *Всероссийская конференция «Информационные технологии интеллектуальной поддержки принятия решений».* 2017:89-96.
7. Splunk и Tensorflow: поиск мошенника с помощью биометрического анализа поведения [Электронный ресурс]. URL: https://www.volgablob.ru/blog/?p=858 (дата обращения: 11.04.2020).
8. Способ и система выявления удаленного подключения при работе на страницах веб-ресурса [Электронный ресурс]. URL: https://edrid.ru/rid/218.016.43e3.html (дата обращения: 11.04.2020).
9. Рублев Д.П., Федоров В.М. Идентификация пользователя по динамическим характеристикам работы с манипулятором «мышь» с использованием нейронных сетей. *Известия ЮФУ. Технические науки*. 2017: 67-71.
10. Система и способ обнаружения приложения удаленного администрирования [Электронный ресурс]. URL: https://edrid.ru/rid/218.016.120b.html (дата обращения: 11.04.2020).
11. Sapozhnikova M.Y., Gayanova M.M., Vulfin A.M., Chuykov A.V., Nikonov A.V. Processing of big data in the transaction monitoring systems. *Труды IV Международной конференции «Информационные технологии и нанотехнологи».* 2018:2526-2533.
12. Звезда И.И. К вопросу о классификации способов мошенничества в банковской сфере. *Известия Тульского государственного университета. Экономические и юридические науки.* 2015;3-2:97-104.
13. Sapozhnikova M.U., Gayanova M.M., Vulfin A.M, Nikonov A.V., Mironov K.V. Data mining technologies in the problem of designing the bank transaction monitoring system. *Computer Science and Information Technologies (CSIT'2017).* 2017:74-84.
14. Sapozhnikova M.U., Gayanova M.M., Vulfin A.M, Nikonov A.V., Mironov K.V., Kurennov D.V. Anti-fraud system on the basis of data mining technologies. *2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE*. 2017:243-248.
15. Никонов А.В., Вульфин А.М., Гаянова М.М., Сапожникова М.Ю. Алгоритмы интеллектуального анализа данных банковских транзакций в составе системы

противодействия финансовому мошенничеству. *Системная инженерия и информационные технологии*. 2019;1:32-40.
16. Sapozhnikova M. U., Nikonov A. V., Vulfin A. M. Intrusion Detection System Based on Data Mining Technics for Industrial Networks. *2018 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). IEEE*. 2018:1-5.
17. Анти-фрод системы и как они работают [Электронный ресурс]. URL: https://www.securitylab.ru/blog/personal/Informacionnaya_bezopasnost_v_detalyah/339929.php (дата обращения: 11.04.2020).
18. Patidar R., Sharma L. Credit Card Fraud Detection Using Neural Network. *International Journal of Soft Computing and Engineering*. 2011;1:32-38.
19. West J., Bhattacharya M. Some Experimental Issues in Financial Fraud Mining. *Procedia Computer Science*. 2016;80:1734-1744.
20. Patel S., Gond S. Supervised Machine (SVM) Learning for Credit Card Fraud Detection. *International Journal of Distributed and Parallel Systems*. 2014;8:137-139.
21. Bhusari V., Patil S. International Journal of Engineering Trends and Technology. *International Journal of Distributed and Parallel Systems*. 2011;2:203-211.
22. Prakash A., Chandrasekar C. An Optimized Multiple Semi-Hidden Markov Model for Credit Card Fraud Detection. *Indian Journal of Science and Technology*. 2015;8:11-18.
23. Matheswaran P., Siva E., Rajesh R. Fraud Detection in Credit Card Using Data Mining Techniques. *International Journal of Distributed and Parallel Systems*. 2015;2:26-34.
24. Huang R., Tawfik H., Nagar A.K. A novel Hybrid Artificial Immune Inspired Approach for Online Break-in Fraud Detection. *Procedia Computer Science*. 2012:2733-2742.
25. Schaidnagel M., Petrov I., Laux F. An Online Algorithm for Credit Card Fraud Detection for Games Merchants. *The Second International Conference on Data Analytics*. 2013:1-6.
26. Patil S., Somavanshi H., Gaikward J., Deshmane A. Credit Card Fraud Detection Using Decision Tree Induction Algorithm. *International Journal of Computer Science and Mobile Computing*. 2015;4:92-95.
27. Real time credit card fraud detection with Apache Spark and Event Streaming [Электронный ресурс]. URL: https://mapr.com/blog/real-time-credit-card-fraud-detection-apache-spark-and-event-streaming/ (дата обращения: 11.04.2020).
28. Real time fraud detection with sequence mining [Электронный ресурс]. URL: https://pkghosh.wordpress.com/2013/10/21/real-time-fraud-detection-with-sequence-mining/ (дата обращения: 11.04.2020).
29. Abbad M., Abed J.M., Abbad M. The Development of E-Banking in Developing Countries in the Middle East. *Journal Financial Account Managemaent*. 2012;2:107-123.
30. Jarrett J.E. Internet Banking Development. *J. Entrep. Organ. Manag*. 2016;5:2-5.
31. Bahnsen A.C., Aouada D., Stojanovic A., Ottersten B. Detecting Credit Card Fraud using Periodic Features. *Computer Science*. 2015;3:37-43.

REFERENCES

1. Russian Foreign Ministry: damage to the global economy from cybercrime in 2019 could reach $ 2 trillion. Available at: https://tass.ru/politika/5551244 (accessed 11.04.2020). (In Russ)
2. Bank card fraud. Available at: http://www.tadviser.ru/index.php/Статья:Мошенничество_с_банковскими_картами (accessed 11.04.2020). (In Russ)
3. Detecting a remote control session using keyboard monitoring methods. Available at: http://www.frodex.ru/article/radkl2015 (accessed 11.04.2020). (In Russ)

4. Remote access to a computer and how to organize it: expanding your business horizons. Available at: https://www.kp.ru/guide/udalennyi-dostup-k-komp-juteru.html (accessed 11.04.2020). (In Russ)
5. Kaspersky Fraud Prevention: A solution for effective protection against cyber fraud. Available at: https://www.karma-group.ru/catalog/kaspersky-for-enterprise/fraud-prevention/ (accessed 11.04.2020). (In Russ)
6. Sapozhnikova M.Y., Vulfin A.M., Gayanova M.M., Nikonov A.V. Data mining algorithms of bank transactions data as a part anti-fraud system. *"Information Technologies for Intellectual Decision Support".* 2017:89-96. (In Russ)
7. Splunk and Tensorflow: Scam Finder Using Biometric Behavior Analysis. Available at: https://www.volgablob.ru/blog/?p=858 (accessed 11.04.2020). (In Russ)
8. Method and system for detecting remote connection when working on web resource pages. Available at: https://edrid.ru/rid/218.016.43e3.html (accessed 11.04.2020). (In Russ)
9. Rublev D.P., Fedorov V.M. Identification of user based on work dynamics with "mouse" pointing device using the neural networks. *Izvestiya SFedU. Engineering Sciences*. 2017: 67-71. (In Russ)
10. System and method for detecting remote administration applications. Available at: https://edrid.ru/rid/218.016.120b.html (accessed 11.04.2020). (In Russ)
11. Sapozhnikova M.Y., Gayanova M.M., Vulfin A.M., Chuykov A.V., Nikonov A.V. Processing of big data in the transaction monitoring systems. *The IV International Conference on Information Technology and Nanotechnology.* 2018:2526-2533.
12. Zvezda I.I. On the classification of fraund in the banking sector. *"Izvestiya Tula State University". Economic and legal sciences.* 2015;3-2:97-104.
13. Sapozhnikova M.U., Gayanova M.M., Vulfin A.M, Nikonov A.V., Mironov K.V. Data mining technologies in the problem of designing the bank transaction monitoring system. *Computer Science and Information Technologies (CSIT'2017).* 2017:74-84.
14. Sapozhnikova M.U., Gayanova M.M., Vulfin A.M, Nikonov A.V., Mironov K.V., Kurennov D.V. Anti-fraud system on the basis of data mining technologies. *2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. *IEEE*. 2017:243-248.
15. Nikonov A.V., Vulfin A.M., Gayanova M.M., Sapozhnikova M.U. Data mining algorithms of bank transactions data as a part anti-fraud system. *SIIT*. 2019;1:32-40. (In Russ)
16. Sapozhnikova M. U., Nikonov A. V., Vulfin A. M. Intrusion Detection System Based on Data Mining Technics for Industrial Networks. *2018 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). IEEE*. 2018:1-5.
17. Anti-fraud systems and how they work. Available at: https://www.securitylab.ru/blog/personal/Informacionnaya_bezopasnost_v_detalyah/339929.php (accessed 11.04.2020). (In Russ)
18. Patidar R., Sharma L. Credit Card Fraud Detection Using Neural Network. *International Journal of Soft Computing and Engineering*. 2011;1:32-38.
19. West J., Bhattacharya M. Some Experimental Issues in Financial Fraud Mining. *Procedia Computer Science*. 2016;80:1734-1744.
20. Patel S., Gond S. Supervised Machine (SVM) Learning for Credit Card Fraud Detection. *International Journal of Distributed and Parallel Systems*. 2014;8:137-139.
21. Bhusari V., Patil S. International Journal of Engineering Trends and Technology. *International Journal of Distributed and Parallel Systems*. 2011;2:203-211.
22. Prakash A., Chandrasekar C. An Optimized Multiple Semi-Hidden Markov Model for Credit Card Fraud Detection. *Indian Journal of Science and Technology*. 2015;8:11-18.

23. Matheswaran P., Siva E., Rajesh R. Fraud Detection in Credit Card Using Data Mining Techniques. *International Journal of Distributed and Parallel Systems*. 2015;2:26-34.
24. Huang R., Tawfik H., Nagar A.K. A novel Hybrid Artificial Immune Inspired Approach for Online Break-in Fraud Detection. *Procedia Computer Science*. 2012:2733-2742.
25. Schaidnagel M., Petrov I., Laux F. An Online Algorithm for Credit Card Fraud Detection for Games Merchants. *The Second International Conference on Data Analytics*. 2013:1-6.
26. Patil S., Somavanshi H., Gaikward J., Deshmane A. Credit Card Fraud Detection Using Decision Tree Induction Algorithm. *International Journal of Computer Science and Mobile Computing*. 2015;4:92-95.
27. Real time credit card fraud detection with Apache Spark and Event Streaming. Available at: https://mapr.com/blog/real-time-credit-card-fraud-detection-apache-spark-and-event-streaming/ (accessed 11.04.2020).
28. Real time fraud detection with sequence mining. Available at: https://pkghosh.wordpress.com/2013/10/21/real-time-fraud-detection-with-sequence-mining/ (accessed 11.04.2020).
29. Abbad M., Abed J.M., Abbad M. The Development of E-Banking in Developing Countries in the Middle East. *Journal Financial Account Managemaent*. 2012;2:107-123.
30. Jarrett J.E. Internet Banking Development. *J. Entrep. Organ. Manag*. 2016;5:2-5.
31. Bahnsen A.C., Aouada D., Stojanovic A., Ottersten B. Detecting Credit Card Fraud using Periodic Features. *Computer Science*. 2015;3:37-43.

ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

Вульфин Алексей Михайлович, доцент кафедры вычислительной техники и защиты информации Уфимского государственного авиационного технического университета, Уфа, Российская Федерация.
e-mail: vulfin.alexey@gmail.com
ORCID: 0000-0002-9358-0651

Aleksey M. Vulfin, Associate Professor, Department of Computer Engineering and Information Protection, Ufa State Aviation Technical University, Ufa, Russian Federation.