Экспериментальная психология
2018. Т. 11. № 3. С. 19—35
doi: 10.17759/exppsy.2018110302
ISSN: 2072-7593
ISSN: 2311-7036 (online)

Experimental Psychology (Russia)
2018, vol. 11, no. 3, pp. 19—35
doi: 10.17759/exppsy.2018110302
ISSN: 2072-7593
ISSN: 2311-7036 (online)

КОЛИЧЕСТВЕННЫЕ КРИТЕРИИ ДЛЯ РАСПОЗНАВАНИЯ НЕКОРРЕКТНОГО ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ КОМПЬЮТЕРНЫХ СЕТЕЙ

КУРАВСКИЙ Л.С.*, *ФГБОУ ВО МГППУ, Москва, Россия,*
e-mail: l.s.kuravsky@gmail.com

ЮРЬЕВ Г.А.**, *ФГБОУ ВО МГППУ, Москва, Россия,*
e-mail: g.a.yuryev@gmail.com

СКРИБЦОВ П.В.***, *компания «Павлин Техно» Москва, Россия,*
e-mail: pvs@pawlin.ru

ЧЕРВОНЕНКИС М.А.****, *компания «Павлин Техно» Москва, Россия,*
e-mail: chervonenkis@yandex.ru

КОНСТАНТИНОВСКИЙ А.А.*****, *ФГБОУ ВО МГППУ, Москва, Россия,*
e-mail: sanekkonst@gmail.com

ШЕВЧЕНКО А.А.******, *ФГБОУ ВО МГППУ, Москва, Россия,*
e-mail: apokend@gmail.com

ИСАКОВ С.С.*******, *ФГБОУ ВО МГППУ, Москва, Россия,*
e-mail: phebra@yandex.ru

Представлены два критерия для выявления отклонений в поведении пользователей при диагностике сетевых угроз. Первый из них опирается на технику проверки статистических гипотез и использует в качестве инструмента для формирования целевой статистики самоорганизующиеся карты признаков (сети Кохонена), представляющие один из видов самообучающихся нейронных сетей. Второй критерий определяет категории пользователей с отклонениями в поведении по выполненным последовательностям типовых действий, используя для представления динамики их поведения марковские процессы с дискретными состояниями и дискретным временем (цепи Маркова).

Для цитаты:
Куравский Л.С., Юрьев Г.А., Скрибцов П.В., Червоненкис М.А., Константиновский А.А., Шевченко А.А., Исаков С.С. Количественные критерии для распознавания некорректного поведения пользователей компьютерных сетей // Экспериментальная психология. 2018. Т. 11. №. 3. С. 19—35. doi:10.17759/exppsy.2018110302

* *Куравский Л.С.* Доктор технических наук, профессор, декан факультета информационных технологий, ФГБОУ ВО МГППУ. E-mail: l.s.kuravsky@gmail.com
** *Юрьев Г.А.* Кандидат физико-математических наук, зам. декана, доцент, факультет информационных технологий, ФГБОУ ВО МГППУ. E-mail: g.a.yuryev@gmail.com
*** *Скрибцов П.В.* Кандидат технических наук, генеральный директор компании «Павлин Техно». E-mail: pvs@pawlin.ru
**** *Червоненкис М.А.* Ведущий разработчик компании «Павлин Техно». E-mail: chervonenkis@yandex.ru
***** *Константиновский А.А.* Студент факультета информационных технологий ФГБОУ ВО МГППУ. E-mail: sanekkonst@gmail.com
****** *Шевченко А.А.* Магистрант факультета информационных технологий ФГБОУ ВО МГППУ. E-mail: apokend@gmail.com
******* *Исаков С.С.* Магистрант факультета информационных технологий ФГБОУ ВО МГППУ. E-mail: phebra@yandex.ru

Ключевые слова: сетевые угрозы, поведение пользователей компьютерных сетей, самоорганизующиеся карты признаков, нейронные сети, цепи Маркова.

Введение

Защита от сетевых угроз в настоящее время является одной из важнейших проблем информационной безопасности компьютерных систем. Применяемые в облачной среде стандартные средства ее поддержки, включая средства идентификации пользователей, ограничения прав доступа и объемов трафика, шифрование данных, программно-аппаратная защита низкого уровня и привлечение в особых случаях операторов в режиме ручного управления, не обеспечивают должную эффективность.

Практический опыт сопровождения компьютерных сетей выявил перспективность выявления возможных угроз на основе анализа поведения пользователей в реальном времени. В частности, компания *Symantec* применяет облачный сервис «*Cloud Access Security Broker*» (*CASB*) (https://www.symantec.com/content/dam/symantec/docs/solution-briefs/secure-use-of-cloud-apps-and-services.pdf), в котором для каждой процедуры, выполняемой пользователем в облаке, методами машинного обучения определяется уровень риска, на основе которого программируется определенный тип поддержки безопасности.

Система *LANeye* (*LANeye Network Intrusion Detection and Prevention Software*) (http://www.laneye.com/software/laneye-product-description.pdf) анализирует трафик пользователя по детерминированным правилам, не применяя методы машинного обучения и сравнивая значения наблюдаемых параметров с аналогичными показателями прошлой сессии.

В системе *UEBA* (*User and Entity Behavior Analytics*), разработанной компанией *Exabeam* (https://www.exabeam.com/data-science/user-entity-behavior-analytics-scoring-system-explained/), применен комплексный метод выявления угроз от пользователей и аномалий в их поведении. Для этого строится набор различных индикаторов, основанных на статистическом анализе, предупреждениях о наличии вредоносных программ, а также на методах машинного обучения (таких как обнаружение доменов *DGA — Domain Generation Algorithm* — с помощью нейросетевых и других способов моделирования). Текстовые данные о пользователе переводятся в числовые с уменьшением размерности с помощью сингулярного разложения (https://www.exabeam.com/data-science/a-user-and-entity-behavior-analytics-system-explained-part-ii/), после чего полученные компактные данные классифицируются методом опорных векторов *SVM* (*Support Vector Machine*). Оценка пользователя формируется как сумма полученных индикаторов с динамически настраиваемыми весами.

Одной из наиболее актуальных научных задач, возникающих при создании подобных систем, является разработка современного математического аппарата для распознавания некорректного поведения пользователей компьютерных сетей, адаптированного к анализу данных, характеризующих сетевую активность, и пригодного для использования в рамках интеллектуальных систем для прогнозирования и выявления угроз. Подобные системы должны работать в облачной среде в автоматическом режиме и, по возможности, обладать способностью к самообучению.

К настоящему времени накоплен определенный опыт в решении этой задачи. Как средство ее решения, специалистами применялись многие хорошо известные методы классификации, включая:

— распознавание с помощью бинарных деревьев решений (Фаткиева, Левоневский, 2015; AlGhamdi et al., 2008);

— динамические и многопользовательские байесовские сети (Дайнеко, 2013; AlGhamdi et al., 2008);

— искусственные нейронные сети (Большев, 2011);

— анализ временных рядов (Фаткиева, 2012; Фаткиева, Левоневский, 2013);

— использование простейших статистических характеристик (Фаткиева, 2012);

— методы анализа графов;

— метод опорных векторов (Mingyuan et al., 2015);

— скрытые марковские модели (Banafar et al., 2014; Hong et al., 2015; Modi, Quadir, 2014);

— генетические алгоритмы (Hameed, 2014; Singh et al., 2016);

— ограниченные машины Больцмана (Hua et al., 2017);

— рассуждения по прецедентам (*case-based reasoning*) (Herrero et al., 2009; Wang et al., 2011);

— методы многомерного статистического анализа, включая кластерный и дискриминантный анализ.

Все они — за исключением классических методов многомерного статистического анализа и различных вариантов использования простейших статистических характеристик — в большей или меньшей степени продемонстрировали свою эффективность, однако общим слабым местом остается отсутствие неэвристических количественных критериев для обоснованного отнесения пользователей к проблемной категории. Указанные выше статистические методы в рассматриваемой предметной области, как правило, дают неприемлемые результаты (см. иллюстрацию их применения в разделе 2.1).

С целью преодоления возникшей проблемы, в этой работе предложены два подхода к распознаванию некорректного поведения пользователей компьютерных сетей, опирающиеся на:

— критерий для выявления отклонений в поведении пользователей: по характеристикам, усредненным на временных интервалах без учета содержательной динамики поведения;

— критерий для определения категорий пользователей по выполненным последовательностям типовых действий (т. е. с учетом содержательной динамики поведения).

Третий подход данного типа — *метод паттернов,* использующий возможности вейвлет-преобразований для диагностики по тестовым траекториям, — представлен в работах (Куравский и др., 2018; Kuravsky, Yuryev, 2018).

Следует отметить, что предложенные методы являются средствами решения достаточно широкого класса задач психологической диагностики и педагогических измерений. В частности, они могут применяться для выявления определенных особенностей в поведении пользователей в социальных сетях, характеризующих психологическое неблагополучие. Критерий для определения категорий пользователей по выполненным последовательностям типовых действий (т. е. с учетом содержательной динамики поведения) найдет свое применение в педагогических измерениях, например, для анализа действий при подготовке к ЕГЭ с целью определения наиболее комфортной траектории обучения. Кроме того, рассмотренные средства имеют хорошие перспективы в исследованиях, связанных с психологией труда, включая оценку психологической усталости оператора сложных систем по изменившимся последовательностям типовых действий.

Количественный критерий для выявления отклонений в поведении пользователей по характеристикам, усредненным на временных интервалах без учета содержательной динамики поведения

Общее описание

Критерий для выявления отклонений в поведении пользователей при диагностике сетевых угроз опирается на применение самоорганизующихся карт признаков (*Self-Organizing Feature Maps*), или *сетей Кохонена* (Kohonen, 2001). Входной слой сети выполняет распределительные функции. На этот слой подаются закодированные (в том числе, если необходимо, используя схему «*Один-из-N*») интегральные характеристики деятельности пользователя за определённые периоды времени, состав и содержание которых могут меняться в зависимости от конкретной решаемой задачи. Выходной слой (топологическая карта) образует прямоугольную матрицу, составленную из элементов на радиальных базисных функциях. При последовательной обработке каждого обучающего примера выбирается расположенный ближе всего к нему нейрон («выигравший» нейрон). Затем, взяв взвешенную сумму прежнего центра соответствующего радиального элемента и обучающего примера, параметры выигравшего нейрона и нейронов из его окрестности корректируются так, чтобы они стали в большей степени похожи на входной пример. Окрестность в процессе обучения сжимается до нулевого отклонения от «выигравшего» нейрона. Результатом последовательности таких корректировок является то, что определенные участки сети «перетягиваются» в сторону обучающих примеров и похожие наблюдения активируют группы близко лежащих нейронов на топологической карте.

Для заданных категорий пользователей (в первую очередь, для пользователей, деятельность которых не представляет опасности для системы) вычисляются выборочные распределения расстояний до выигравшего нейрона. При этом предполагается, что пользователи с отклонениями в поведении присутствуют в обучающей выборке в определенной небольшой пропорции, не оказывая существенного влияния на результат обучения. Пользователи с относительно редким поведением фактически рассматриваются как потенциально опасные. Важно, что данное предположение позволяет не выполнять предварительное распознавание пользователей с отклонениями в поведении в исходных эмпирических данных. Если пользователь с относительно редким поведением рассматривается как неопасный, то его следует включить в обучающее множество. Если пользователь с опасным поведением похож на представителей «неопасных» классов, то его следует исключить из обучающего множества.

Полученные выборочные распределения в дальнейшем используются для проверки статистических гипотез о принадлежности пользователей к заданным классам. При этом в качестве статистики, для которой вычисляются вероятности, сопоставляемые с уровнем значимости, используется расстояние до выигравшего нейрона. Уровень значимости является параметром постановки задачи. Его стандартное значение — *0,05*, однако, в зависимости от содержания прикладной задачи, этот показатель может варьироваться от *0,01* до *0,1*.

Представленная технология распознавания типов пользователей представлена на рис. 1, где в условии $p<p*$ использованы следующие обозначения: $p*$ — уровень значимости для проверки гипотезы, $p=1-F(X)$, $X = \min_{i \in I} r(N_i)$, $r(N_i)$ — евклидово расстояние от набора характеристик деятельности оцениваемого пользователя до нейрона N_i сети Кохонена, $i \in I$ — индекс нейрона, I — множество индексов нейронов, $F(X)$ — выборочная функция распределения случайной величины X.

Если гипотезы о принадлежности к «безопасным» классам отвергаются при принятом уровне значимости ($p<p*$) или, при том же уровне значимости, при наличии соответствующих эмпирических данных не отвергаются гипотезы о принадлежности к «опасным» классам ($p\geq p*$), то пользователь идентифицируется как представляющий опасность.

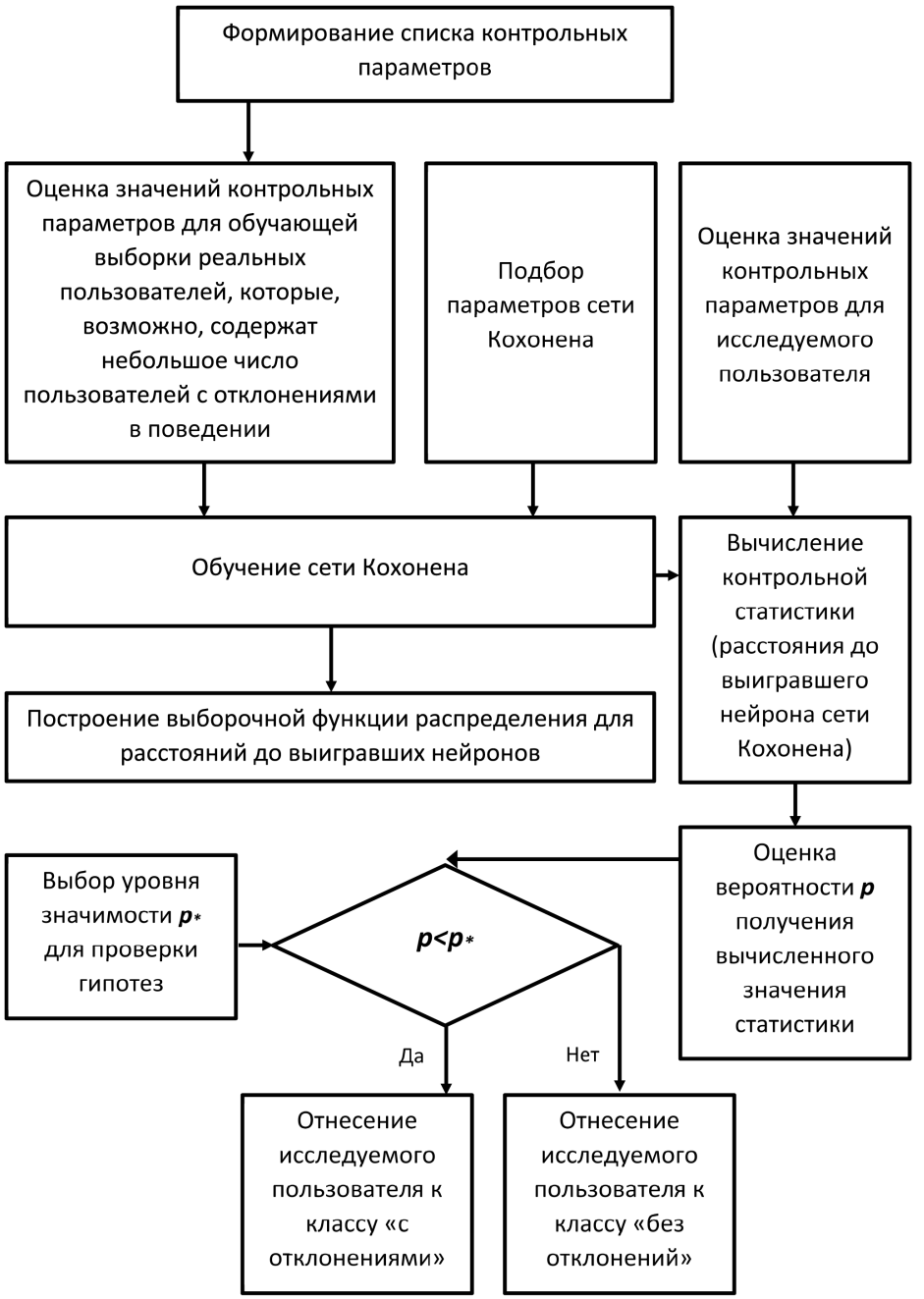

Рис. 1. Технология распознавания типов пользователей

Новизна рассмотренного подхода заключается в том, что:

— для формирования статистики, используемой для проверки гипотез о принадлежности к выявляемым классам пользователей, используются сети Кохонена, представляющие один из видов самообучающихся нейронных сетей;

— вычисленные с их помощью выборочные распределения используются для оценки вероятностей, сопоставляемых с уровнем значимости.

Пример построения критерия

Для построения критерия использовалась сформированная путем эксперимента обучающая выборка из 323 пользователей, 318 из которых принадлежали к 3 классам с «безопасным» поведением (*«programmer»*, *«serfer»* и *«lazyman»*), а 5 — к классу пользователей с отклонениями в поведении (*«violator»*). Показатели пользовательской активности, на основе которых строились оценки, представлены в отчете (см. Отчет о прикладных научных исследованиях и экспериментальных разработках на тему «Разработка интеллектуальных алгоритмов выявления сетевых угроз в облачной вычислительной среде и методов защиты от них, основанных на анализе динамики трафика и определении отклонений в поведении пользователей». Этап 1. ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технического комплекса России на 2014—2020 годы»).

Топологическая карта сети, размеченная по этой выборке, представлена на рис. 2. Выборочная плотность вероятности и выборочная функция распределения расстояний до «выигравших» нейронов представлены, соответственно, на рис. 3 и 4.

Для оценки надежности распознавания использовалась контрольная выборка из 55 пользователей с отклонениями в поведении, параметры которых были выявлены в процессе экспериментов. Оценки расстояний до «выигравшего» нейрона для элементов указанной выборки позволили вычислить выборочную плотность распределения, показанную на рис. 5. Минимальное расстояние до «выигравшего» нейрона при этом составило 0,34,

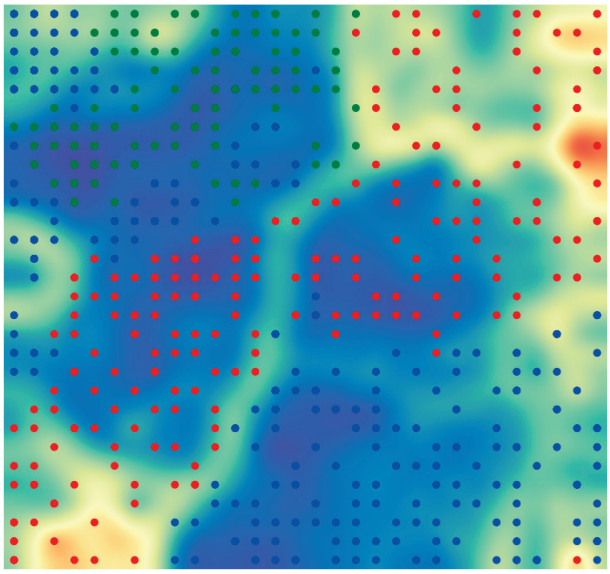

Рис. 2. Топологическая карта сети Кохонена, размеченная по обучающей выборке

Рис. 3. Выборочная плотность вероятности расстояний до «выигравших» нейронов

Рис. 4. Выборочная функция распределения расстояний до «выигравших» нейронов

максимальное — 1,88. Выборочная функция распределения для обучающей выборки, представленная на рис. 3, позволяет утверждать, что вероятность появления расстояний до «выигравшего» нейрона, превышающих минимальное расстояние, равное 0,34, в случае пользователей без отклонений в поведении не превышает 0,015.

Поэтому проверки нулевых гипотез о том, что пользователи из контрольной выборки (с отклонениями в поведении) относятся к «безопасным» классам, привели к тому, что указанные нулевые гипотезы были отвергнуты при высоком уровне значимости ($p<0,015$), и все пользователи с отклонениями были правильно идентифицированы как не относящиеся к «безопасным» классам.

Рис. 5. Выборочная плотность вероятности расстояний до «выигравших» нейронов для элементов контрольной выборки, состоящей из пользователей с отклонениями в поведении

Имеющиеся экспериментальные данные свидетельствуют о высокой надежности распознавания как вследствие высокого уровня значимости при отвержении гипотез, так и вследствие полного отсутствия ошибок при распознавании пользователей с отклонениями для контрольной выборки. Таким образом, *предложенный критерий для распознавания пользователей с отклонениями продемонстрировал высокую эффективность на доступных эмпирических данных*.

Распределения значений рассматриваемых параметров не позволяют применить для классификации пользователей классический дискриминантный анализ вследствие отклонений от нормальности и статистически значимых отличий матриц ковариаций для разных типов пользователей, однако этот метод можно использовать для грубой оценки степени их дискриминации.

Статистика Уилкса для полного набора из 48 параметров составляет 0,18 ($F(144,966)=5,1204$; $p<0,0001$), что свидетельствует о статистически значимой, но относительно грубой дискриминации. *Дискриминантный анализа Фишера*[1] обеспечил 75%-е распознавание типов пользователей, при этом «опасные» пользователи распознавались только в 47,3% случаев, что не является удовлетворительным результатом. Матрица классификации приведена в табл. 1.

Удаление 17 переменных, которые не значимы для распознавания типов пользователей, методом «*Forward Stepwise*» повысило значение статистики Уилкса до 0,20 ($F(93,1015)=7,6180$; $p<0,0001$), снизив процент корректного распознавания до 73,5%, при этом процент распознавания «опасных» пользователей уменьшился до 41,8% (см. матрицу классификации в табл. 2).

На рис. 6—8 приведены *диаграммы рассеяния*, качественно иллюстрирующие дискриминацию рассматриваемых типов пользователей при *каноническом дискриминантном*

[1] Для вычислений использовался пакет статистического анализа *STATISTICA*.

Таблица 1

Матрица классификации в случае 48 параметров

	Percent	programmer	serfer	violater	lazyman
programmer	90.0	180	16	3	1
serfer	52.6	25	50	20	0
violater	47.3	20	9	26	0
lazyman	100.0	0	0	0	23
Total	74.8	225	75	49	24

Таблица 2

Матрица классификации в случае 31 параметра

	Percent	programmer	serfer	violater	lazyman
programmer	90.0	180	15	4	1
serfer	50.5	29	48	18	0
violater	41.8	21	11	23	0
lazyman	100.0	0	0	0	23
Total	73.5	230	74	45	24

анализе в собственном подпространстве, базис которого задает направления наибольшей неоднородности обучающей совокупности наблюдений. Для формирования указанного собственного подпространства выбираются собственные вектора, которые соответствуют первым по порядку наибольшим собственным значениям, объясняющим достаточно высокий процент наблюдаемой дисперсии.

Характеристики собственного подпространства, использованного для построения диаграмм рассеяния, приведены в табл. 3.

Таблица 3

Характеристики собственного подпространства, использованного для построения диаграмм рассеяния

No	Eigenvalue	Wilks' Lambda	Chi-Square	df	p-level
0	1.49	0.20	566.93	93	0.000
1	0.80	0.50	243.25	60	0.000
2	0.10	0.91	34.58	29	0.219

Хорошо видно, что пользователи с опасным поведением не отделяются от остальных групп. Неудовлетворительный процент распознавания «опасных» пользователей и качественный анализ взаимного расположения пользователей различных типов в рассмотренном собственном подпространстве позволяют говорить о невозможности распознавания «опасных» пользователей с помощью классических методов дискриминантного анализа. В то же время критерий, опирающийся на возможности сетей Кохонена, эффективно решает эту задачу.

Таким, образом, можно утверждать, что *предложенный критерий демонстрирует существенно более высокую эффективность, чем классические методы многомерного статистического анализа.*

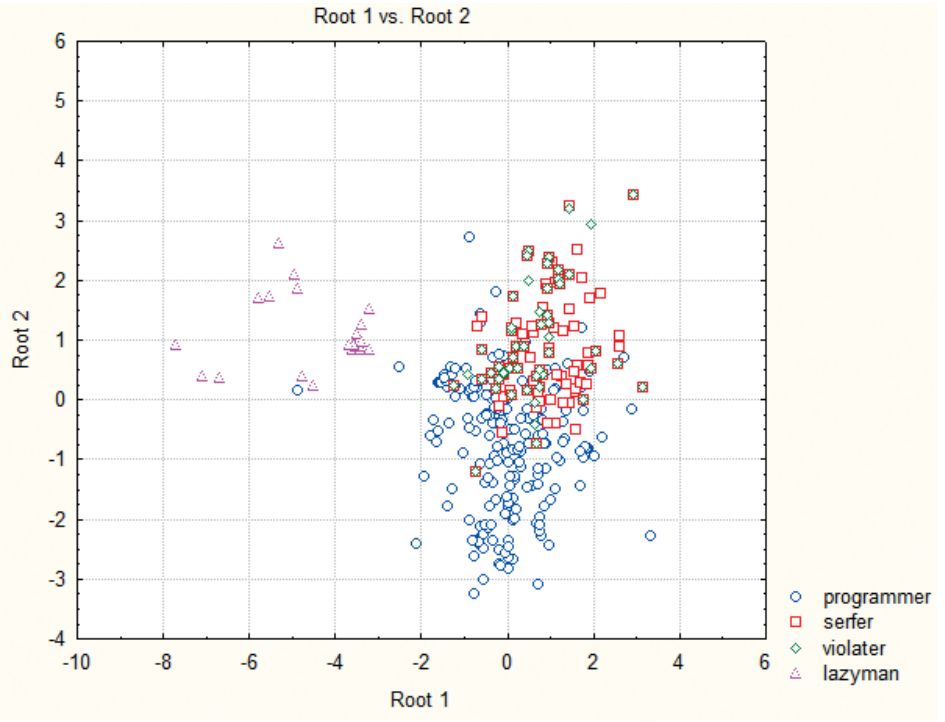

Рис. 6. Диаграмма рассеяния в пространстве собственных осей 1—2

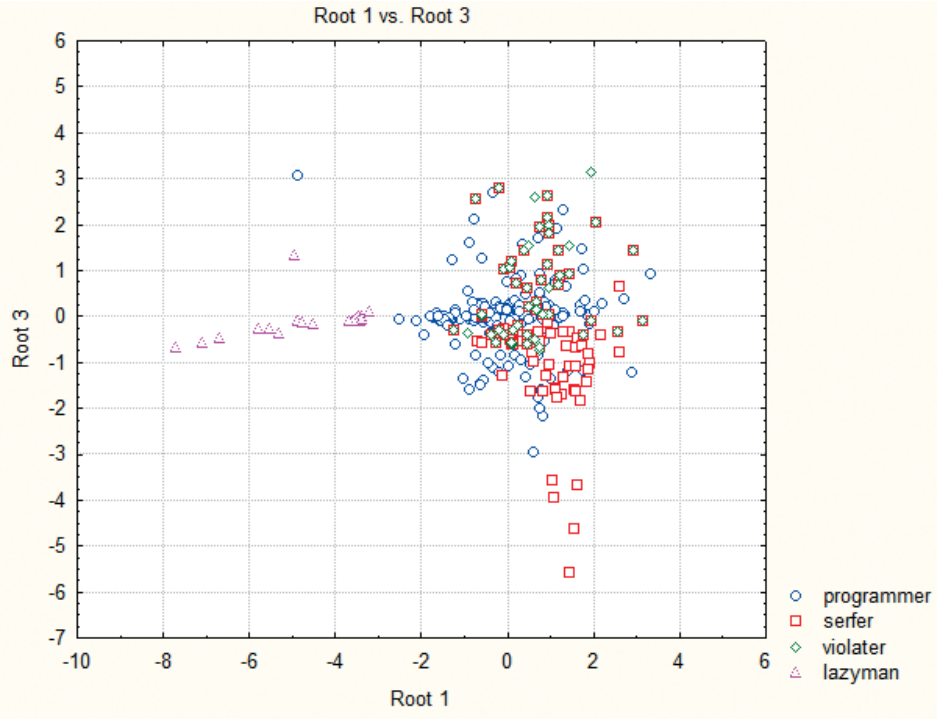

Рис. 7. Диаграмма рассеяния в пространстве собственных осей 1—3

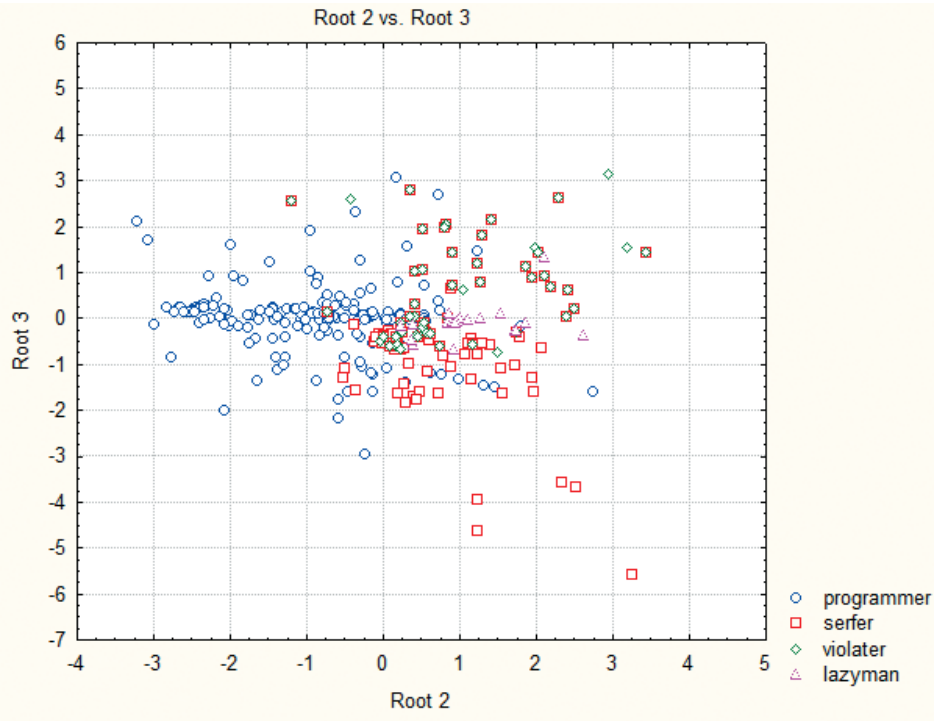

Рис. 8. Диаграмма рассеяния в пространстве собственных осей 2—3

Количественный критерий для определения категорий пользователей по выполненным последовательностям типовых действий с учетом содержательной динамики поведения

Для представления динамики поведения пользователей используются *марковские процессы с дискретными состояниями и дискретным временем (цепи Маркова)*. В этих моделях типовым действиям пользователей (таким как открытие, копирование, удаление, пересылка файлов, имеющих заданные форматы и диапазоны размеров, запуск определенных типов приложений и т. д.) соответствуют определенные состояния, а вероятности переходов между состояниями являются параметрами модели и определяются типом пользователя. Каждой категории пользователей $l \in \{0, ..., z\}$, включая пользователей как с корректным, так и некорректным поведением, соответствует своя модель с уникальным набором вероятностей переходов между состояниями.

Поведение пользователей характеризуется последовательностями выполненных ими типовых действий, которые в терминах данной модели интерпретируются как последовательности состояний.

Динамика вероятностей пребывания в состояниях модели как функций дискретного времени определяется следующим матричным уравнением:

$$\boldsymbol{p}\,(t+1) = \mathbf{M}_l \mathbf{p}(t),$$

где t — дискретное время; $0 \le t \le T$; $t, T \in N$; T — конечный момент времени; N — множество натуральных чисел; $\mathbf{p}(t) = (p_0(t), \ldots, p_n(t))^T$ — представляет вероятности пребывания в состояниях модели в момент времени t; n — число состояний; $\mathbf{M}_l = \| m_{ij,l} \|$ — стохастическая квадратная матрица вероятностей перехода между состояниями цепи Маркова порядка n,

в которой $m_{ij,i}$ — вероятность перехода из состояния j в состояние i для пользователя категории l.

Идентификация рассмотренных марковских моделей выполняется, используя эмпирические данные о частотах переходов от одного типового действия к другому для каждой рассматриваемой категории пользователей. Каждая категория пользователей l имеет свою идентифицированную матрицу $\mathbf{M}_l$.

Отнесение пользователей к одной из заданных категорий $l \in \{0, ..., z\}$ выполняется на основе выполненных им типовых действий, заданных последовательностью пройденных состояний $S_r = \{s_1, s_2, ... , s_r\}$. При этом для каждой из указанных категорий вычисляется соответствующая байесовская оценка:

$$P(C_l|S) = \frac{P(C_l)P(S|C_l)}{\sum_{k=0}^{z} P(C_k)P(S|C_k)},$$

где C_l — факт принадлежности пользователя к категории l; S — событие, представляющее собой прохождение последовательности состояний S_r; $P(C_l)$ — априорная вероятность принадлежности пользователя к категории l; $P(S|C_l)$ — вероятность прохождения последовательности состояний S_r при условии принадлежности к категории l; $P(C_l|S)$ — вероятность принадлежности к категории l при условии, что пользователь прошел последовательность состояний S_r.

Для вычисления вероятностей используются элементы матриц $\mathbf{M}_l$:

$$P(S|C_l) = \prod_{k=1}^{r-1} m_{s_{k+1}s_k,l}$$

Категория пользователей, для которой достигается максимальная условная вероятность $P(C_{max}|S) = \max_l\{P(C_l|S)\}_{l=0,\dots,z}$, обеспечивает требуемый выбор. Распределение вероятностей $\{P(C_l|S)\}_{l=0,\dots,z}$ позволяет оценить его надежность.

Примеры практического применения критериев данного типа представлены в работах (Куравский и др., 2016; Куравский, Юрьев, 2011; 2012; Куравский и др., 2017; 2018; Kuravsky et al., 2016).

Основные выводы и результаты

1. Разработан критерий для выявления отклонений в поведении пользователей при диагностике сетевых угроз, опирающийся на технику проверки статистических гипотез и использующий в качестве инструмента для формирования целевой статистики сети Кохонена, представляющие один из видов самообучающихся нейронных сетей. Особенностями подхода являются:

— оценка вероятностей, сопоставляемых с уровнем значимости, непосредственно по выборочным распределениям расстояний до выигравшего нейрона, полученным для обучающей выборки, без построения аналитического выражения целевой статистики;

— возможность обучения сети Кохонена на смешанной выборке, допускающей наличие в определенной небольшой пропорции потенциально опасных пользователей, что позволяет избежать необходимости их выявления на ранних этапах исследования, когда не известны соответствующие идентифицирующие признаки.

2. Предварительная оценка, проведенная с использованием доступных экспериментальных данных, выявила высокую эффективность предложенного подхода: для потенциально опасных пользователей гипотеза об их принадлежности к «безопасным» классам отвергалась при уровне значимости не более 0,015; все 100% потенциально опасных пользователей были распознаны. Применение классических методов многомерного статистического анализа, выполненное для сравнения различных подходов на тех же данных, выявило, что пользователи с опасным поведением не отделяются от остальных групп классическими способами. В частности, неудовлетворительный процент распознавания (<50%) опасных пользователей и качественный анализ взаимного расположения пользователей различных типов в рассмотренном собственном подпространстве позволили говорить о невозможности распознавания этой категории пользователей с помощью классических методов дискриминантного анализа.

3. Разработан метод определения категорий пользователей, включая пользователей с отклонениями в поведении, по выполненным последовательностям типовых действий, использующий для представления динамики поведения пользователей марковские процессы с дискретными состояниями и дискретным временем (цепи Маркова). Особенностями подхода являются:

— представление поведения пользователей последовательностями выполненных ими типовых действий, которые в терминах применяемой модели интерпретируются как последовательности состояний;

— использование для каждой категории пользователей, включая пользователей как с корректным, так и некорректным поведением, отдельной модели с уникальным набором вероятностей переходов между состояниями;

— отнесение пользователей к одной из заданных категорий на основе байесовских оценок и оценок правдоподобия.

Финансирование

Работа выполнена при финансовой поддержке Министерства образования и науки Российской федерации в рамках соглашения о предоставлении субсидии от «26» сентября 2017 г. № 14.579.21.0155 (Уникальный идентификатор соглашения — RFMEFI57917X0155) на выполнение прикладных научных исследований и экспериментальных разработок по теме: «Разработка интеллектуальных алгоритмов выявления сетевых угроз в облачной вычислительной среде и методов защиты от них, основанных на анализе динамики трафика и определении отклонений в поведении пользователей».

Литература

1. *Большев А.К.* Алгоритмы преобразования и классификации трафика для обнаружения вторжений в компьютерные сети: дисс. ... канд. техн. наук: 05.13.11, 05.13.19 СПб.: Гос. электротехн. ун-т (ЛЭТИ), 2011. 155 с.

2. *Дайнеко В.Ю.* Разработка модели и алгоритмов обнаружения вторжений на основе динамических байесовских сетей: дисс. ... канд. техн. наук: 05.13.19. СПб.: Нац. исслед. ун-т информац. технологий, механики и оптики, 2013. 131 с.

3. *Куравский Л.С., Марголис А.А., Мармалюк П.А., Панфилова А.С., Юрьев Г.А.* Математические аспекты концепции адаптивного тренажера // Психологическая наука и образование. 2016. Т. 21. № 2. С. 84—95. doi: 10.17759/pse.2016210210

4. *Куравский Л.С., Юрьев Г.А.* Вероятностный метод фильтрации артефактов при адаптивном тестировании // Экспериментальная психология. 2012. Т. 5. № 1. С. 119—131.

5. *Куравский Л.С., Юрьев Г.А.* Использование марковских моделей при обработке результатов тестирования // Вопросы психологии. 2011. № 2. С. 98—107.

6. *Куравский Л.С., Юрьев Г.А., Ушаков Д.В., Поминов Д.А., Юрьева Н.Е., Валуева Е.А., Лаптева Е.М.* Диагностика по тестовым траекториям: метод паттернов // Экспериментальная психология. 2018. Т. 11. №. 2. С. 77—94. doi:10.17759/exppsy.2018110206.
7. Марковские модели в задачах диагностики и прогнозирования: Учеб. пособие / Под ред. Л.С. Куравского. 2-е изд., доп. М.: Изд-во МГППУ, 2017. 203 с.
8. Отчет о прикладных научных исследованиях и экспериментальных разработках на тему «Разработка интеллектуальных алгоритмов выявления сетевых угроз в облачной вычислительной среде и методов защиты от них, основанных на анализе динамики трафика и определении отклонений в поведении пользователей» // № госрегистрации АААА-А17-117122890077-5. Этап 1. ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технического комплекса России на 2014—2020 годы». Соглашение о предоставлении субсидии № 14.579.21.0155 от 26.09.2017.
9. *Фаткиева Р.Р.* Корреляционный анализ аномального сетевого трафика // Труды СПИИРАН. 2012. Вып. 23. С. 93—99.
10. *Фаткиева Р.Р.* Модель обнаружения атак на основе анализа временных рядов // Труды СПИИРАН. 2012. Вып. 2. С. 71—80.
11. *Фаткиева Р.Р., Левоневский Д.К.* Детектирование компьютерных атак методом сингулярного спектрального разложения // Труды СПИИРАН, 2013. Вып. 25. С. 135—147.
12. *Фаткиева Р.Р., Левоневский Д.К.* Применение бинарных деревьев для агрегации событий систем обнаружения вторжений // Труды СПИИРАН. 2015. Вып. 40. С. 110—121.
13. «CatchSync»: Catching Synchronized Behavior in Large Directed Graphs. [Электронный ресурс]. URL: http://www.meng-jiang.com/pubs/catchsync-kdd14/catchsync-kdd14-paper.pdf (дата обращения: 09.02.2018).
14. *AlGhamdi G.A., Laskey K.B., Wright E.J., Barbara D., and Chang K.* Modeling insider user behavior using multi-entity Bayesian network // 10th International Command and Control Research and Technology Symposium. 2008. Vol. 4444. № 703.
15. *Banafar H., Sharma, S.* Intrusion Detection and Prevention System for Cloud Simulation Environment using Hidden Markov Model and MD5 // International Journal of Computer Applications. 2014. Vol. 90. № 19. P. 6—11. doi: 10.5120/15826-4490
16. *Hameed U.N., Ahamd F., Alyas T., Khan, W.* Intrusion Detection and Prevention in Cloud Computing using Genetic Algorithm // International Journal of Scientific and Engineering Research. 2014. Vol. 5.
17. *Herrero A, Corchado E. In: Abraham A, Hassanien A-E, de Carvalho A, Editors.* Mining Network Traffic Data for Attacks through MOVICAB-IDS Foundations of Computational Intelligence, 4 204. Berlin/ Heidelberg: Springer; 2009. Pp. 377—94
18. *Hong B., Peng F., Deng B., Hu Y., Wang D.* DAC-Hmm: detecting anomaly in cloud systems with hidden Markov models // Concurrency Computat.: Pract. Exper. 2015. Vol: 27. Pp. 5749—5764. doi: 10.1002/ cpe.3640
19. *Hua Zhang, Shixiang Zhu, Xiao Ma, Jun Zhao, Zeng Shou.* A Novel RNN-GBRBM Based Feature Decoder for Anomaly Detection Technology in Industrial Control Network. IEICE Transactions. 2017. Pp. 1780—1789.
20. *Kohonen T.* Self-Organizing Maps, Springer. 3th Ed. 2001. 501 p.
21. *Kuravsky L.S., Marmalyuk P.A., Yuryev G.A., Belyaeva O.B., Prokopieva O.Yu.* Mathematical Foundations of Flight Crew Diagnostics Based on Videooculography Data [Электронный ресурс] // Applied Mathematical Sciences. 2016. Vol. 10. № 30 P. 1449—1466. URL: http://dx.doi.org/10.12988/ ams.2016.6122.
22. *Kuravsky L.S., Yuryev G.A.* On the approaches to assessing the skills of operators of complex technical systems. // In: Proc. 15th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Nottingham, UK. 2018. 25 pp.
23. *Modi K., Quadir A.* Detection and Prevention of DDoS Attacks on the Cloud using Double-TCP Mechanism and HMM-based Architecture // International Journal of Cloud Computing and Services Science (IJ-CLOSER). 2014. Vol. 3.
24. Secure use of cloud apps & services. CABS. Cloud Access Security Broker. Symantec [Электронный ресурс]. URL: https://www.symantec.com/content/dam/symantec/docs/solution-briefs/secure-use-of-cloud-apps-and-services.pdf (дата обращения: 09.02.2018).

25. *Singh T., Verma S., Kulshrestha V., Katiyar S.* Intrusion Detection System Using Genetic Algorithm for Cloud. // In: Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (ICTCS '16). ACM, New York, NY, USA. 2016. Article 115. 6 pages. DOI: http://dx.doi.org/10.1145/2905055.2905175
26. *Wang Y., Anguo Z.,Jichun Z.* A Case-Based Reasoning Method for Network Security Situation Analysis. International Conference on Control, Automation and Systems Engineering (CASE). 2011. Pp. 1—4.
27. *Yu M., Huang S., YuQ., Wang Y., Gao J.* A Density-based Binary SVM Algorithm in the Cloud Security // International Journal of Security and Its Applications. 2015. Vol. 9. Pp. 153—162. doi: 10.14257/ijsia.2015.9.7.14

QUANTITATIVE CRITERIA FOR RECOGNIZING THE INCORRECT BEHAVIOR OF COMPUTER NETWORK USERS

KURAVSKY L.S.*, *Moscow State University of Psychology & Education, Moscow, Russia, e-mail: l.s.kuravsky@gmail.com*

YURYEV G.A.**, *Moscow State University of Psychology & Education, Moscow, Russia, e-mail: g.a.yuryev@gmail.com*

SCRIBTSOV P.V.***, *"Pavlin-Techno", Moscow, Russia, e-mail: pvs@pawlin.ru*

CHERVONENKIS M.A.****, *"Pavlin-Techno", Moscow, Russia, e-mail: chervonenkis@yandex.ru*

KONSTANTINOVSKY A.A.*****, *Moscow State University of Psychology & Education, Moscow, Russia, e-mail: sanekkonst@gmail.com*

SHEVCHENKO A.A.******, *Moscow State University of Psychology & Education, Moscow, Russia, e-mail: apokend@gmail.com*

ISAKOV S.S.*******, *Moscow State University of Psychology & Education, Moscow, Russia, e-mail: phebra@yandex.ru*

Two approaches for recognizing the incorrect behavior of computer network users are presented. The first one relies on the technique of statistical hypotheses testing and uses self-organizing feature maps (Ko-

For citation:
Kuravsky L.S., Yuryev G.A., Scribtsov P.V., Chervonenkis M.A., Konstantinovsky A.A., Shevchenko A.A., Isakov S.S. Quantitative criteria for recognizing the incorrect behavior of computer network users. *Eksperimental'naya psikhologiya= Experimental psychology (Russia)*, 2018, vol. 11, no. 3, pp. 19—35. doi:10.17759/exppsy.2018110302

* *Kuravsky L.S.* PhD, Professor, Moscow State University of Psychology & Education. E-mail: l.s.kuravsky@gmail.com
** *Yuryev G.A.* PhD, Docent (Associate Professor), Moscow State University of Psychology & Education. E-mail: g.a.yuryev@gmail.com
*** *Scribtsov P.V.* PhD, CEO « Pavlin Techno ». E-mail: pvs@pawlin.ru
**** *Chervonenkis M.A.* Leading researcher, « Pavlin Techno ». E-mail: chervonenkis@yandex.ru
***** *Konstantinovsky A.A.* Student, Moscow State University of Psychology & Education. E-mail: sanekkonst@gmail.com
****** *Shevchenko A.A.* Master student, Moscow State University of Psychology & Education. E-mail: apokend@gmail.com
******* *Isakov S.S.* Master student, Moscow State University of Psychology & Education. E-mail: phebra@yandex.ru

honen networks) for generating target statistics. The second approach recognizes dangerous activity using executed sequences of relevant typical actions, with their dynamics being represented with the aid of Markov chains.

Keywords: computer network threats, user activity, self-organizing feature maps, Markov chains.

Funding

The study was supported by the Russian Ministry of Education and Science, № 14.579.21.0155.

References

1. Bol'shev A.K. *Algoritmy preobrazovaniya i klassifikacii trafika dlya obnaruzheniya vtorzhenij v komp'yuternye seti: diss. ... kand. tekhn. Nauk [Algorithms of classification of traffic for inclusion in computer networks. PhD thesis].* 05.13.11, 05.13.19 SPb, Gos. ehlektrotekhn. un-t (LEHTI), 2011, 155 p. (In Russ.).
2. Dajneko V.YU. *Razrabotka modeli i algoritmov obnaruzheniya vtorzhenij na osnove dinamicheskih bajesovskih setej: diss. ... kand. tekhn. Nauk [Development of a model and algorithms of detection of inclusions based on dynamic Bayesian networks. PhD thesis].* SPb, Nac. issled. un-t informac. tekhnologij, mekhaniki i optiki, 2013, 131 p. (In Russ.).
3. Kuravskiy L.S., Margolis A.A., Marmalyuk P.A., Panfilova A.S., YUr'ev G.A. Matematicheskie aspekty koncepcii adaptivnogo trenazhera [Mathematical aspects of the conception of an adaptive training]. *Psihologicheskaya nauka i obrazovanie [Psychological science and education],* 2016, vol. 21, no. 2, pp. 84—95. doi: 10.17759/pse.2016210210. (In Russ.).
4. Kuravskiy L.S., Yuriev G.A. Veroyatnostnyj metod fil'traciiartefaktov pri adaptivnom testirovanii [Probability method of filtration in adaptive testing]. *Eksperimental'naya psihologiya [Experimental Psychology],* 2012, vol. 5. no. 1, pp. 119—131. (In Russ.).
5. Kuravskiy L.S., Yuriev G.A. Ispol'zovanie markovskih modelej pri obrabotke rezul'tatov testirovaniya [Using Markov models for testing analysis]. *Voprosy psihologii [Issues in Psychology],* 2011, no. 2, pp. 98—107.
6. Kuravskiy L.S., Yuriev G.A., Ushakov D.V., Pominov D.A., Yurieva N.E., Valueva E.A., Lapteva E.M. Diagnostika po testovym traektoriyam: metod patternov [Diagnostic of testing trajectories: method of patterns]. *Eksperimental'naya psihologiya [Experimental Psychology],* 2018, vol. 11, no. 2, pp. 77—94. doi:10.17759/exppsy.2018110206. (In Russ.).
7. *Markovskie modeli v zadachah diagnostiki i prognozirovaniya: Ucheb. Posobie [Markov models in diagnostics and prognosis. Manuel].* L.S. Kuravskoy (Eds.). Moscow, Izd-vo MGPPU, 2017, 203 p. (In Russ.).
8. Otchet o prikladnyh nauchnyh issledovaniyah i ehksperimental'nyh razrabotkah na temu «Razrabotka intellektual'nyh algoritmov vyyavleniya setevyh ugroz v oblachnoj vychislitel'noj srede i metodov zashchity ot nih, osnovannyh na analize dinamiki trafika i opredelenii otklonenij v povedenii pol'zovatelej» // № gosregistracii AAAA-A17-117122890077-5. Etap 1. FCP «Issledovaniya i razrabotki po prioritetnym napravleniyam razvitiya nauchno-tekhnicheskogo kompleksa Rossii na 2014—2020 gody». Soglashenie o predostavlenii subsidii № 14.579.21.0155 ot 26.09.2017. (In Russ.).
9. Fatkieva R.R. Korrelyacionnyj analiz anomal'nogo setevogo trafika [Correlation analysis of abnormal internet traffic]. *Trudy SPIIRAN*, 2012, no. 23, pp. 93—99. (In Russ.).
10. Fatkieva R.R. Model' obnaruzheniya atak na osnove analiza vremennyh ryadov [Model of detection of attacks based on time analysis]. *Trudy SPIIRAN*, 2012, no. 2, pp. 71—80. (In Russ.).
11. Fatkieva R.R., Levonevskij D.K. Detektirovanie komp'yuternyh atak metodom singulyarnogo spektral'nogo razlozheniya [Detecting of computer attacks using singular spectral method]. *Trudy SPIIRAN*, 2013, no. 25, pp. 135—147. (In Russ.).
12. Fatkieva R.R., Levonevskij D.K. Primenenie binarnyh derev'ev dlya agregacii sobytij sistem obnaruzheniya vtorzhenij [Using binary trees for agregations of events in systems of inclusion detecting]. *Trudy SPIIRAN*, 2015, no. 40, pp. 110—121. (In Russ.).
13. «CatchSync»: Catching Synchronized Behavior in Large Directed Graphs. URL: http://www.mengjiang.com/pubs/catchsync-kdd14/catchsync-kdd14-paper.pdf

14. AlGhamdi G.A., Laskey K.B., Wright E.J., Barbara D., and Chang K. Modeling insider user behavior using multi-entity Bayesian network. *10th International Command and Control Research and Technology Symposium,* 2008, vol. 4444, no. 703.
15. Banafar H., Sharma, S. Intrusion Detection and Prevention System for Cloud Simulation Environment using Hidden Markov Model and MD5. International *Journal of Computer Applications*, 2014, vol. 90, no. 19, pp. 6—11. doi: 10.5120/15826-4490
16. Hameed U.N., Ahamd F., Alyas T., Khan, W. Intrusion Detection and Prevention in Cloud Computing using Genetic Algorithm. *International Journal of Scientific and Engineering Research,* 2014, vol. 5.
17. Herrero A, Corchado E. In: Abraham A, Hassanien A-E, de Carvalho A, Editors. *Mining Network Traffic Data for Attacks through MOVICAB-IDS Foundations of Computational Intelligence*, 4 204. Berlin Heidelberg, Springer, 2009, pp. 377—94
18. Hong B., Peng F., Deng B., Hu Y., Wang D. DAC-Hmm: detecting anomaly in cloud systems with hidden Markov models. *Concurrency Computat, Pract. Exper*, 2015, vol. 27, pp. 5749—5764. doi: 10.1002/cpe.3640
19. Hua Zhang, Shixiang Zhu, Xiao Ma, Jun Zhao, Zeng Shou. A Novel RNN-GBRBM Based Feature Decoder for Anomaly Detection Technology in Industrial Control Network. *IEICE Transactions*, 2017, pp. 1780—1789.
20. Kohonen T. *Self-Organizing Maps*. Springer. 2001, 501 p.
21. Kuravsky L.S., Marmalyuk P.A., Yuryev G.A., Belyaeva O.B., Prokopieva O.Yu. Mathematical Foundations of Flight Crew Diagnostics Based on Videooculography Data. *Applied Mathematical Sciences*, 2016, vol. 10, no. 30, pp. 1449—1466. URL: http://dx.doi.org/10.12988/ams.2016.6122.
22. Kuravsky L.S., Yuryev G.A. On the approaches to assessing the skills of operators of complex technical systems. In Proc. *15th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies*, Nottingham, UK, 2018, 25 pp.
23. Modi K., Quadir A. Detection and Prevention of DDoS Attacks on the Cloud using Double-TCP Mechanism and HMM-based Architecture. *International Journal of Cloud Computing and Services Science (IJ-CLOSER),* 2014, vol. 3.
24. Secure use of cloud apps & services. CABS. Cloud Access Security Broker. Symantec. URL: https://www.symantec.com/content/dam/symantec/docs/solution-briefs/secure-use-of-cloud-apps-and-services.pdf
25. Singh T., Verma S., Kulshrestha V., Katiyar S. Intrusion Detection System Using Genetic Algorithm for Cloud. In: *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*. New York, NY, USA, 2016, Article 115, 6 pages. DOI: http://dx.doi.org/10.1145/2905055.2905175
26. Wang Y., Anguo Z., Jichun Z. A Case-Based Reasoning Method for Network Security Situation Analysis. *International Conference on Control, Automation and Systems Engineering (CASE),* 2011, pp. 1—4.
27. Yu M., Huang S., Yu Q., Wang Y., Gao J. A Density-based Binary SVM Algorithm in the Cloud Security. *International Journal of Security and Its Applications,* 2015, vol. 9, pp. 153—162. doi: 10.14257/ijsia.2015.9.7.14