**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

|  |  |
| --- | --- |
|  | **ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ**  **Кафедра**  **«Криптология и кибербезопасность»** |

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Отчет**

**о научно-исследовательской работе**

«Обнаружение внутреннего нарушителя путём выявления стрессового состояния пользователя на основе анализа взаимодействия с клавиатурой и мышью с применением алгоритмов машинного обучения»

|  |  |  |
| --- | --- | --- |
| Исполнитель:  студент гр. Б16-506 | \_\_\_\_\_\_\_\_\_\_\_ (подпись, дата) | Султанов А.Э. |
| Научный руководитель: | \_\_\_\_\_\_\_\_\_\_\_ (подпись, дата) | Когос К.Г. |
| Зам. зав. каф. № 42: | \_\_\_\_\_\_\_\_\_\_\_ (подпись, дата) | Когос К.Г. |

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Москва – 2019**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**РЕФЕРАТ**

СОДЕРЖАНИЕ

**ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ**

# В настоящем отчете применяются следующие термины с соответствующими определениями, обозначениями и сокращениями:

# RF — метод «случайного леса» (Random Forest).

# k-NN — метод k-ближайших соседей (k-nearest neighbors algorithm).

SVM — метод опорных векторов (Support Vector Machine).

GB — метод градиентного бустинга (Gradient Boosting).

DT — дерево принятия решений (Decision Tree).

LR — логистическая регрессия (Logistic Regression).

ROC — рабочая характеристика приёмника (Receiver Operating Characteristic).

AUC — площадь под ROC-кривой (Area Under ROC curve).

DBN — сеть глубоких убеждений (Deep Belief Network).

GSS — метод золотого сечения (Golden Section Search).

IF — изолирующий лес (Isolation Forest).

PCA — метод главных компонент (Principal Component Analysis).

DNN — глубокая нейронная сеть (Deep Neural Network).

RNN — рекуррентная нейронная сеть (Recurrent Neural Network).

BA — бэггинг или бутстрэп-агрегирование (Bootstrap aggregating).

# ВВЕДЕНИЕ

В современном мире информационных технологий обеспечение информационной безопасности становится всё более сложной задачей. Несмотря на огромные усилия со стороны компаний, разрабатывающих средства эффективного противодействия информационным угрозам, для некоторых из угроз данная задача остаётся всё ещё нерешённой.

Одной из таких угроз, с которой сталкиваются большинство компаний, в той или иной степени связанных с миром технологий, является угроза внутреннего нарушителя. Данный вид угрозы представляет наибольшую опасность в силу огромного числа порождающих её источников и недостатка эффективных средств противодействия этой угрозе, вследствие чего компании несут огромные финансовые потери. Именно поэтому угроза внутреннего нарушителя является актуальной и на данный момент ей уделяется значительное внимание со стороны исследователей в мире информационной безопасности.

Объяснением отсутствия достаточного количества средств борьбы против угрозы внутреннего нарушителя является характер данной угрозы. Несмотря на то, что результат действий внутреннего нарушителя очевиден, очень сложно найти информативные показатели, которые позволили бы обнаружить аномальное поведение пользователя информационной системы и отличить его от нормального. Для анализа поведения можно использовать данные электрокардиограммы (ЭКГ) или электроэнцефалограммы (ЭЭГ) [1,2,3], температуры тела [1], проводимости кожи[2], движения глаз [4] и других биометрических показателей.

Однако способы получения большинства видов биометрических показателей являются инвазивными, то есть требуют использования специального и дорогостоящего оборудования в виде датчиков и камер, которые чаще всего находятся в непосредственном контакте с участником эксперимента в процессе накопления данных. Этот недостаток является причиной того, что методы, разработанные на основе инвазивных способов получения биометрических показателей, сложно применить на практике.

При этом существуют некоторые неинвазивные способы накопления данных для анализа поведения, которые основы на использовании доступных инструментов, например клавиатуры и мыши, которые легко можно найти в любом офисе. Эти недорогостоящие инструменты могут выступать в качестве датчиков, которые предоставляют поведенческую информацию или, иными словами, поведенческую характеристику, состоящую из клавиатурного почерка, динамики нажатия клавиш и жестов. Поведенческую характеристику можно использовать в решении задач аутентификации [5,6,7,8,9], детекции эмоционального состояния [10,11,12,13,14,15], а также в задачах обнаружения угрозы внутреннего нарушителя [16].

Данная работа посвящена исследованию методов обнаружения внутреннего нарушителя путём выявления стрессового состояния пользователя на основе анализа взаимодействия с клавиатурой и мышью. Как и в работе [16], исследование построено на проверке предположения о том, что при совершении неправомерных действий у внутреннего нарушителя под воздействием индуцированного стрессового состояния меняются поведенческие показатели. Однако в отличие от работы [16] применяется более обширный диапазон выделенных из сырых данных признаков и для подтверждения гипотезы используется большее число алгоритмов.

В первом разделе проведён анализ сопутствующих работ и выявлены недостатки, связанные с используемыми данными и методами их обработки, а также приведены результаты классификации применённых алгоритмов.

Во втором разделе рассмотрен процесс накопления данных, включающий в себя требования к сценариям для накопления данных, их описание и обоснование использования данных сценариев.

Во третьем разделе описан процесс выделения признаков из сырых данных, их обработка, а также рассмотрены модели классификаторов аномального и нормального поведения, построенные на основе различных алгоритмов машинного обучения.

В четвёртом разделе приведены результаты применённых моделей и проведена сравнительная характеристика этих моделей.

1. **Сопутствующие работы**

Угрозу внутреннего нарушителя могут представлять текущий или бывший работник, сотрудник или бизнес-партнёр, которые имеют или имели привилегии доступа к сети, системе или данным организации и преднамеренно или непреднамеренно выполняют действия, которые негативно влияют на конфиденциальность, целостность и доступность организации или информационной системы [17].

Основными типами внутренних нарушителей являются те, кто действуют преднамеренно и злонамеренно и те, кто представляют угрозу организации в силу своей невнимательности, небрежности и совершения нечаянных действий [18]. К основным целям и задачам первой категории можно отнести:

* саботаж;
* кража интеллектуальной собственности;
* шпионаж;
* мошенничество с целью финансовой выгоды;

Причины угроз со стороны членов второй категории:

* человеческие ошибки;
* фишинг;
* вредоносные программы;
* непреднамеренное пособничество;
* украденные учётные записи;

Угроза внутреннего нарушителя рассматривается в качестве основной проблемы безопасности организаций [19] и по данным, приведённым в работе [20] приблизительно 87% случаев угроз безопасности организации зафиксированы, как угрозы со стороны внутренних нарушителей. В данной работе будут рассмотрены только случаи преднамеренного создания угрозы безопасности внутренним нарушителем и приведены основные методы обнаружения с применением алгоритмов машинного обучения.

Для обнаружения угрозы внутреннего нарушителя применяются две основные категории алгоритмов машинного обучения:

* методы классификации из группы алгоритмов с учителем;
* методы кластеризации из группы алгоритмов без учителя;
  1. **Обнаружение угрозы с применением алгоритмов с учителем**

В работе [21] проведён сравнительный анализ классификаторов, обученных на общедоступном датасете CERT Insider Threat database, являющийся набором большого количества реальных случаев угроз внутреннего нарушителя с информацией о траффике веб-ресурсов и почтовых ресурсов, а также логах файловой системы. В качестве алгоритмов, составляющих основу классификаторов, были выбраны такие, как RF, k-NN, GB, DT, LR, их вариации и сочетания. Лучшие результаты были показаны классификаторами на основе алгоритма RF и его вариаций. Достигнутая точность как на полном, так и на урезанном датасете варьировалась от 94% до 98%. Однако классификаторы на основе RF проиграли в быстроте обучения классификаторам на основе k-NN, которые в лучшем случае выдали точность предсказания немного хуже максимума точности в случае RF.

Аналогичный датасет был проанализирован в работе [22], однако в данной работе помимо алгоритмов RF, SVM, DT и LR были применены нейронные сети. Для каждой из моделей была также создана версия с использованием алгоритмов бустинга, что позволило увеличить точность предсказаний почти для всех моделей. Результаты моделей на основе нейронных сетей не превзошли точности моделей на основе RF, DT и LR, при этом минимальная точность классификации среди всех классификаторов составила почти 92%.

Работа [23] выделяется среди остальных тем, что обучение моделей классификаторов происходит не на статическом датасете, а на непрерывном потоке данных. В качестве алгоритма для модели классификатора используется OCSVM, который после приведения обучаемой выборки в пространство признаков с большой размерностью, рассматривается как обычный SVM. Максимальная полученная точность классификации составила 71%.

В статье [24] предпринята попытка создания модели мультиклассового классификатора на основе алгоритма k-NN для решения задачи двухфакторной аутентификации. Классификатор предсказывал принадлежность пользователя, основываясь на распознавании лица, к одной из четырёх групп:

* разрешённый;
* возможно разрешённый;
* возможно неразрешённый;
* неразрешённый;
  1. **Обнаружение угрозы с применением алгоритмов без учителя**

В работе [25] обнаружение угрозы внутреннего нарушителя реализована с применением предварительной кластеризации данных на основе графов и последующем выделении аномалий. В качестве датасета использована база данных CERT. При оценке модели наилучший результат показателя AUC составил 0.76.

Ещё один метод кластеризации был применён в статье [26]. Модель основана на одной из разновидностей глубоких нейронных сетей DBN. В качестве датасета выбран общедоступный CERT, однако модель обучена только на данных логов поведения пользователей. Сеть DBN была оптимизирована с помощью алгоритма GSS, а результат работы сети был использован для обучения классификатора на основе SVM. В результате полученная точность обнаружения угрозы внутреннего нарушителя составила 97.8%.

В статье [27] авторы представили систему обнаружения угрозы внутреннего нарушителя из больших потоков данных в режиме реального времени, основанную на глубоких и рекуррентных нейронных сетях. Для уменьшения размерности пространства признаков использовался алгоритм PCA. Модели оценивались с помощью IF и SVM. В результате как для модели на основе DNN, так и для модели на основе RNN получен recall в 100%.

* 1. **Обнаружение угрозы путём выявления стрессового состояния**

Существуют работы, в которых обнаружение угрозы внутреннего нарушителя основано на предположении о том, что при совершении аномальных действий пользователь, который является текущим или бывшим сотрудником организации, испытывает стресс. Стресс в свою очередь влияет на такие биометрические показатели как пульс, кровяное давление, температура тела, а также на ритм и динамику нажатия клавиш на устройствах, посредством которых происходит взаимодействие с информационной системой. Следовательно, выявив стресс с использованием биометрических показателей, можно выстроить цепь в обратном порядке и с определённой долей вероятности предполагать, что стресс был вызван из-за неправомерных действий, представляющих собой угрозу внутреннего нарушителя. Такое предположение сделано в работах [1,2,3,4].

В ресурсе [3] проведено комплексное исследование, включающее процессы сбора данных, предобработки, обучения моделей классификаторов и их оценка. В качестве основных биометрических показателей для отслеживания стресса были выбраны сигналы электрокардиограммы (ЭКГ) или электроэнцефалограммы (ЭЭГ). Наибольший интерес представляет процесс сбора данных, в котором используются сценарии, описывающие как действия внутреннего нарушителя, вызывающие стрессовое состояние, так и действия обычного сотрудника компании, не представляющие угрозу безопасности. В качестве основы модели классификатора выбран SVM. В результате, точность предсказания модели классификатора составила 86%.

Большим недостатком с точки зрения практического применения в работе [3] является необходимость наличия специального оборудования, отслеживающего изменения биометрических показателей. Для решения данной проблемы в статье [16] в качестве отслеживающих датчиков используются клавиатура и мышь, которые дают ценную информацию о динамике нажатия клавиш и кнопок. Как уже было показано в работах [10,11,12,13,14,15] клавиатурный почерк меняется под воздействием различных эмоциональных состояний. В работе [16] было построено четыре модели классификаторов на основе SVM, RF, k-NN и BA. Модели были обучены на данных, собранных с помощью двух категорий сценариев, как и в работе [3], описывающих действия внутреннего нарушителя и обычного сотрудника. Полученные точности моделей варьировались от 67.5% до 72.5%. Авторы объясняют плохие результаты недостатком информативных признаков, выделенных из динамики использования клавиатуры. Большинство признаков было выделено из динамики использования мыши, однако в выводах отмечено, что влияние стресса на динамику очевидно, и это подтверждает предположение о том, что умение выявлять стресс посредством биометрических показателей позволяет обнаруживать угрозу внутреннего нарушителя.

Текущее исследование во многом схоже с работой [16], однако присутствуют существенные различия в процессах накопления данных и выделении признаков, а также в использованных алгоритмах машинного обучения для моделей классификаторов.

1. **Накопление данных**
2. **Выделение признаков и модели**
3. **Результаты**

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. **Abdulaziz A.** On the Possibility of Insider Threat Detection Using Physiological Signal Monitoring [Текст] / A. Abdulaziz, E. Khalil // Proc. of the ACM 7th International Conference on Security of Information and Networks — 2014.
2. **Hojae L.** An Application of Data Leakage Prevention System based on Biometrics Signals Recognition Technology [Текст] / L. Hojae, J. Junkwon, K. Taeyoung, P. Minwoo, E. Jungho, T.M. Chung // SUComS — 2013.
3. **Yessir H.** Inside the Mind of the Insider: Towards Insider Threat Detection Using Psychophysiological Signals [Текст] / H. Yessir, T. Hasan, G. Mohammad, D. Ram // Journal of Internet Services and Information Security — 2016.
4. **Hassan T.** Prediction of Human Error Using Eye Movements Patterns for Unintentional Insider Threat Detection [Текст] / T. Hassan, H. Yessir, D. Ram // IEEE 4th International Conference on Identity, Security, and Behavior Analysis — 2018.
5. **Bergadano F.** Identity verification through dynamic keystroke analysis [Текст] / F. Bergadano, D. Gunneti, C. Picardi // Intelligence Data Analisys Journal 7 — 2003.
6. **Dowland P.** A Long-term trial of keystroke profiling using digraph, trigraph and keyword latencies [Текст] / P. Dowland, S. Furnell // In IFIP International Federation for Information Processing Journal — 2004.
7. **Joyce R.** Identity authentication based on keystroke latencies [Текст] / R. Joyce, G. Gupta // Commun. ACM 33 — 1990.
8. **Monrose F.** Keystroke dynamics as a biometric for authentication [Текст] / F. Monrose, A. D. Rubin // Future Generation Computing Systems 16 — 2000.
9. **Bender S.** Key sequence rhythm recognition system and method [Текст] / Bender, S. and Postley, H // (U.S. Patent № 7 206 938) — 2002.
10. **Kołakowska A.** Recognizing emotions on the basis of keystroke dynamics [Текст] / A. Kołakowska // 2015 8th International Conference on Human System Interaction (HSI) — 2015.
11. **Epp C.** Identifying Emotional States using Keystroke Dynamics [Текст] / C. Epp, M. Lippold, R. L. Mandryk // Proceedings of the International Conference on Human Factors in Computing Systems — 2011.
12. **Nazmul Haque F.M.** Identifying Emotion by Keystroke Dynamics And Text Pattern Analysis [Текст] / F.M. Nazmul Haque, J.M. Alam // Behaviour and Information Technology Journal — 2012.
13. **Suranga D.W.G.** Non Invasive Human Stress Detection Using Key Stroke Dynamics and Pattern Variations [Текст] / D.W.G. Surang, M. De Silva Pasan, S.B.K. Dayan, M.K.D. Arunatileka Shiromi // International Conference on Advances in ICT for Emerging Regions (ICTer) — 2013.
14. **Kołakowska A.** Towards detecting programmers’ stress on the basis of keystroke dynamics [Текст] / A. Kołakowska // Proceedings of the Federated Conference on Computer Science and Information Systems — 2016.
15. **Ulinskas M.** Recognition of human daytime fatigue using keystroke data [Текст] / M. Ulinskasa, R. Damaševičiusa, R. Maskeliūnasa, M. Woźniak // The Workshop on Computational Intelligence in Ambient Assisted Living Systems (CIAALS 2018) — 2018.
16. **Yassir H.** Insider Threat Detection Based on Users’ Mouse Movements and Keystrokes Behavior [Текст] / H. Yassir, T. Hassan, D. Ram // Conference Secure Knowledge Management Conference — 2017.
17. **CERT Definition of 'Insider Threat'** [Электронный ресурс] — Режим доступа: https://insights.sei.cmu.edu/insider-threat/2017/03/cert-definition-of-insider-threat---updated.html — Свободный.
18. **Insider Threat** [Электронный ресурс] — Режим доступа: https://www.observeit.com/insider-threat/ — Свободный.
19. **Insider Threats as the Main Security Threat in 2017** [Электронный ресурс] — Режим доступа: https://www.tripwire.com/state-of-security/security-data-protection/insider-threats-main-security-threat-2017/ — Свободный.
20. **Oladimeji T.O.** Review on Insider Threat Detection Techniques [Текст] / T.O. Oladimeji, C.K. Ayo, S.E. Adewumi // 3rd International Conference on Science and Sustainable Development (ICSSD 2019) — 2019.
21. **David A.N** Classifier Suites for Insider Threat Detection [Текст] / A.N.David // Arxiv Journal, Machine Learning (cs.LG) — 2019.
22. **Adam J.H.** Predicting Malicious Insider Threat Scenarios Using Organizational Data and a Heterogeneous Stack-Classifier [Текст] / J.H.Adam, P. Nikolas, J.B. William, M. Naghmeh // Arxiv Journal, Machine Learning (cs.LG) — 2019.
23. **Parveen P.** Supervised Learning for Insider Threat Detection Using Stream Mining [Текст] / P. Parveen, R.Z. Weger, B. Thuraisingham, H. Kevin, K. Latifur // IEEE 23rd International Conference on Tools with Artificial Intelligence — 2019.
24. **Sarma M.S.** Insider Threat Detection with Face Recognition and KNN User Classification [Текст] / M.S. Sarma, Y. Srinivas, M. Abhiram, L. Ullala, M.S. Prasanthi, J.R. Rao // IEEE International Conference on Cloud Computing in Emerging Markets (CCEM) — 2017.
25. **Anagi G.** Insider Threat Detection Through Attributed Graph Clustering [Текст] / G. Anagi, B. Serdar // Proceedings of the 16th IEEE International Conference on Trust, Security and Privacy in Computing and Communications — 2017.
26. **Zhang J.** Insider threat detection of adaptive optimization DBN for behavior logs [Текст] / J. Zhang, Y. Chen, J. Ankang // Turkish Journal of Electrical Engineering & Computer Sciences — 2017.
27. **Choras M.** Machine Learning Techniques for Threat Modeling and Detection [Текст] / M. Choras, R. Kozik // Security and Resilience in Intelligent Data-Centric Systems and Communication Networks — 2018 — С. 179-192.