Исхаков А.Ю., Мамченко М.В.

Алгоритм аутентификации пользователей на основе поведенческой аналитики и машинного обучения для веб-ресурсов

Аннотация: Повышение эффективности аутентификации пользователей на веб-ресурсах продолжает оставаться актуальной задачей в области информационной безопасности. Для решения этой проблемы предлагается алгоритм обнаружения аномалий в поведении пользователей на основе машинного обучения, стандартных журналов (логов) безопасности и цифровых отпечатков браузеров пользователей. Максимальная точность классификации составила 78,3% (IsolationForest). Результаты эксперимента позволяют сделать вывод о возможности использования классификаторов IsolationForest и EllipticEnvelope для решения данной задачи.

Ключевые слова: многофакторная аутентификация, браузерные цифровые отпечатки, обнаружение аномалий, обнаружение выбросов в данных, поведение пользователей, веб-ресурс, машинное обучение

Введение

В настоящее время механизмы аутентификации пользователей являются одним из обязательных средств обеспечения информационной безопасности веб-ориентированных систем, в том числе киберфизических (далее – веб-ресурсов) и объектов критической информационной инфраструктуры. Чаще всего подразумевают именно многофакторную аутентификацию (МФА; совместное использование паролей, PIN-кодов и прочих механизмов аутентификации) [1-3].

С другой стороны, применение МФА является затратным с точки зрения вычислительных ресурсов, особенно в случае наличия большого числа пользователей [4]. Для снижения вычислительных затрат могут применяться интеллектуальные методы и алгоритмы, например, анализ поведения пользователей, обнаружение

поведенческих аномалий и адаптивные методы аутентификации [5-10].

1. Алгоритм обнаружения аномалий в поведении пользователей

В работе предлагается алгоритм для выявления аномалий поведения пользователей с использованием машинного обучения. Для каждого пользователя формируется отдельный классификатор (вместо использования общего классификатора или обобщенных данных о поведении всех пользователей), при этом в ходе обучения возможно применение нескольких классификаторов и выбор рационального среди них (наиболее подходящего к конкретной задаче) [3].

Нормальное поведение для каждого пользователя определяется при обучении модели на основе записей о его действиях. В случае обычного поведения действия пользователя разрешаются, данные об этих действиях добавляются к набору данных для дальнейшего обучения. При обнаружении аномального поведения используются один или несколько факторов аутентификации для проверки личности пользователя. В случае недостоверности действия пользователя запрещаются. Схема алгоритма представлена на рисунке 1.

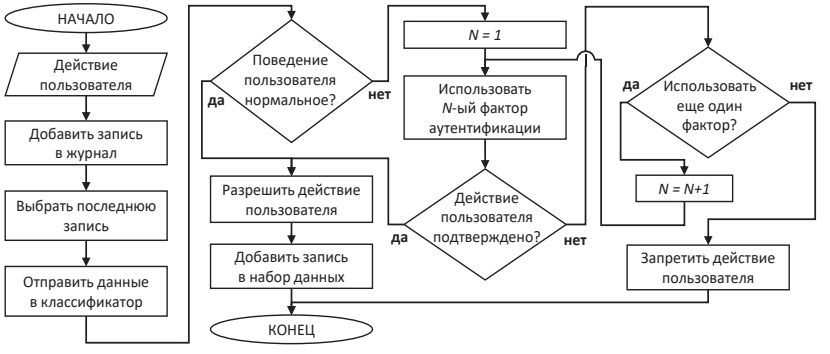

Рисунок 1 – Схема предлагаемого алгоритма обнаружения аномалий поведения пользователей

2. Эксперимент

В эксперименте в качестве набора данных использовался стандартный журнал аудита с 48229 записями и девятью признаками. Затем набор признаков был расширен за счет использования данных из цифровых браузерных отпечатков пользователей (с применением библиотеки FingerprintJS) [3, 8].

Для обнаружения аномалий в действиях пользователей были применены классификаторы OneClassSVM, EllipticEnvelope и IsolationForest.

В таблице 1 представлены значения матрицы ошибок для всех трех классификаторов. OneClassSVM дает большое значение FN (False Negative), т.е. обычное поведение ошибочно классифицируется как аномальное. В этом случае МФА будет задействоваться часто, в том числе при отсутствии угроз безопасности. Для классификатора IsolationForest средняя точность возрастает, для пользователей с небольшим объемом данных значение FN стремится к нулю, а значение FP (False Positive; аномальное поведение ошибочно классифицируется как обычное) увеличивается. Для классификатора EllipticEnvelope значения матрицы ошибок аналогичны результатам применения IsolationForest, при этом средняя точность ниже.

Таблица 1 – Значения матрицы ошибок для трех классификаторов

Количество записей для пользователя	**True Positive**	**True Negative**	**False Positive**	**False Negative**
OneClassSVM				
485 записей	198	8	0	279
80 записей	40	4	0	36
14 записей	3	9	0	2
IsolationForest				
485 записей	381	5	2	97
80 записей	76	3	1	0
14 записей	8	1	5	0
EllipticEnvelope				
485 записей	410	3	2	70
80 записей	69	8	3	0
14 записей	8	3	3	0

3. Выбор рационального классификатора

Для вышеуказанных трех классификаторов были получены следующие значения средней точности:

OneClassSVM – 75,3%;

IsolationForest – 78,3%;

EllipticEnvelope – 76,6%.

На рисунке 2 показан график зависимости точности классификаторов от количества записей в журнале, используемых в качестве набора данных. Применение IsolationForest дает большую точность для небольшого количества записей, в то время как EllipticEnvelope показывает лучшие результаты для больших объемов данных. Наконец, OneClassSVM показывает самую низкую точность для всех случаев.

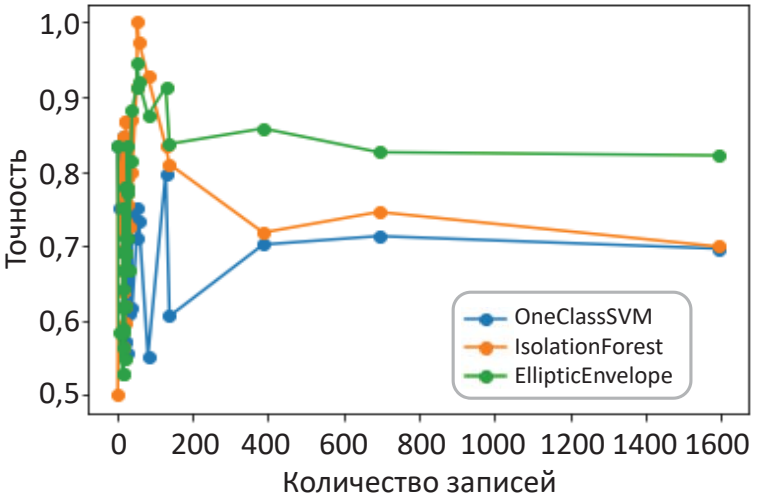

Рисунок 2 – Средняя точность для трех классификаторов

Заключение

Для повышения эффективности аутентификации на веб-ресурсах и выявления аномалий в поведении пользователей в работе предложен соответствующий алгоритм, основанный на машинном обучении. Максимальная точность классификации составила 78,3% (IsolationForest). Результаты эксперимента

свидетельствуют о целесообразности использования IsolationForest и EllipticEnvelope в качестве основных классификаторов.

Литература:

1. *Iskhakov A.Y., Iskhakova A.O., Meshcheryakov R.V., Smirnov A.M.* Authentication model for mobile access subjects // IFAC-PapersOnLine. – 2022. – Vol. 55. No. 9. – P. 222-226. – DOI: 10.1016/j.ifacol.2022.07.039.

2. *Iskhakov A., Meshcheryakov R., Okhapkina E.* Method of access subject authentication profile generation / Proceedings 2020 International Russian Automation Conference (RusAutoCon). – Sochi: IEEE, 2020. – P. 431-436. – DOI: 10.1109/RusAutoCon49822.2020.9208222.

3. *Iskhakov A.Y., Salomatin A.A.* Estimation of the time for calculating the attributes of browser fingerprints in the user authentication task / Topical Problems of Agriculture, Civil and Environmental Engineering (TPACEE 2020). – E3S Web of Conferences. – 2020. – Vol. 224. No. 01030. – P. 1-9. – DOI: 10.1051/e3sconf/202022401030.

4. *Yang S., Meng J.* Research on Multi-factor Bidirectional Dynamic Identification Based on SMS / IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). – IEEE: Chongqing, 2018. – P. 1578-1582. – DOI: 10.1109/IAEAC.2018.8577505.

5. *Çetin U., Tasgin M.* Anomaly Detection With Multivariate K-sigma Score Using Monte Carlo / 5th International Conference on Computer Science and Engineering (UBMK). – IEEE: Diyarbakir, 2020. – P. 94-98. – DOI: 10.1109/UBMK50275.2020.9219482.

6. *Wang Y. et al.* Iterative anomaly detection / IEEE International Geoscience and Remote Sensing Symposium (IGARSS). – IEEE: Fort Worth, 2017. – P. 586-589. – DOI: 10.1109/IGARSS.2017.8127021.

7. *Goldman A., Cohen I.* Anomaly detection based on an iterative local statistics approach / 23rd IEEE Convention of Electrical and Electronics Engineers in Israel. – IEEE: Tel-Aviv, 2004. – P. 440-443. – DOI: 10.1109/EEEI.2004.1361186.

8. *Kim G.-Y., Lim S.-M., Euom I.-C.* A Study on Performance Metrics for Anomaly Detection Based on Industrial Control System

Operation Data // Electronics. – 2022. – Vol. 11. No. 8. – 1213. – DOI: 10.3390/electronics11081213.

9. *Zoppi T., Gharib M., Atif M., Bondavalli A.* Meta-Learning to Improve Unsupervised Intrusion Detection in Cyber-Physical Systems // ACM Transactions on Cyber-Physical Systems. – 2021. – Vol. 5. No. 4. – Article No. 42. – P. 1-27. – DOI: 10.1145/3467470.

10.*Zhong M., Zhou Y., Chen G.* A Security Log Analysis Scheme Using Deep Learning Algorithm for IDSs in Social Network // Security and Communication Networks. – 2021. – Vol. 2021. – Article ID 5542543. – P. 1-13. – DOI: 10.1155/2021/5542543.

Уймин А.Г.

Система непрерывно-дискретной биометрической идентификации на основе анализа потока данных компьютерной мыши

Аннотация: В данной работе рассматривается система непрерывно-дискретной биометрической идентификации, использующая данные движения компьютерной мыши для улучшения точности и надежности идентификации пользователей. Введение таких биометрических систем является актуальным в условиях глобализации цифровых технологий, поскольку обеспечивает защиту персональных данных и позволяет создать надежные механизмы аутентификации, учитывающие поведенческие особенности пользователя. Исследование направлено на разработку алгоритмов, способных эффективно анализировать пользовательские паттерны взаимодействия с мышью, что позволит повысить безопасность и адаптивность системы.

Ключевые слова: биометрическая идентификация, движения мыши, безопасность данных, поведенческие паттерны, алгоритмы анализа данных

В эпоху глобализации цифровых технологий и стремительного развития информационного общества, обеспечение личного