Московский государственный университет имени М. В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра автоматизации систем вычислительных комплексов

Калякина Алина Дмитриевна

Исследование и разработка методов обнаружения аномального поведения пользователей на основе анализа содержимого потока обрабатываемых текстовых данных для задач информационной безопасности

КУРСОВАЯ РАБОТА

Научный руководитель:
к. ф.-м. н., ???
Д. В. Царёв

Москва, 2020

Аннотация

Тут будет аннотация

Содержание

Введение

Инсайдер - человек, который в силу своего служебного положения или иных обстоятельств имеет доступ к конфиденциальной информации внутри компании. **Инсайдерская угроза** — это вредоносная для комании угроза, исходящая от инсайдера.

Инсайдеров и инсайдерские атаки можно классифицировать по:

- **Типу доступа**: по сети или физическому. По своей природе инсайдер имеют авторизованный доступ к сети и/или физический доступ к данным компании. Исследования показывают, что 2/3 инцидентов совершены через сеть.[1]

- **Стратегии инсайдера**: предатель, притворщик и непреднамеренный инсайдер. Предатели — пользователи, которые находятся внутри организации и умышленно злоупотрябляют своими правами. Притворщики - внешние злоумышленники, которые используют украденные идентификационные данные чтобы выдать себя за инсайдера. Непреднамеренный инсайдер — текущий сотрудник, который без злого умысла причинает вред или увеличивает уязвимость компании в будущем. Исследования показывают, что 92% исайдерских атак осуществляются именно предателями.[1]

- **Виду атаки**: саботаж, кража (интеллектуальной собственности), мошенничество и шпионаж. Саботаж - нанесение вреда организации, например, заведомое создание “бекдоров” , под мошенничеством чаще вывод средств организации обманным путем. Самый частый вид атаки - саботаж[1], обычно его совершают недовольные работники.

Ущерб от инсайдерских атак растет с каждым годом, так по данным Ponemon[2] суммарный ущерб от подобных атак вырос на 31% с \$8.76 млн в 2018 году до \$11.45 млн в 2020 году. Кроме того на 47% увеличилось количество инцидентов (с 3,200 в 2018 до 4,716 в 2020). По данным

Для снижения ущерба от инсайдерских атак возможно компании используют следующие технологии и практики (перечислены пять наиболее эффективных по данным[2] в порядке убывания эффективности):

1. **UEBA-системы (англ. User and Entity behavior anaytics)** анализаруют поведение пользователей. Более подробно рассмотрим их в дальнейшем.

2. **Системы управления доступом (PAM, англ. Privileged access management)**. Доступ к данным предоставляется только непосредственно использующих их сотрудникам, что уменьшает риск утечек.

3. **Обучение сотрудников**. Многие сотрудники совершают инсайдерские атаки не преднамеренно: редко меняют пароль, случайно разглашают данные, постоянное

обучение помогает предотвратить такие инциденты. По данным Ponemon это самый часто применяемый метод, хотя и не самый эффективный.

4. **SIEM-системы (англ. Security incident&event)** анализируют в реальном времени события безопасности.

5. **Обмен разведданными об угорозах**.

При составлении приведенного выше рейтинга Ponemon учитывали затраты на внездрение практики и ее поддержку. Интересно, DLP-системы (англ. Data loss prevention) вторые по популярности по данныи Ponemon занимают последнее место в рейтинге эффективности. DLP-системы следят за пересекающими периметр информационный системы потоками данных.

Основное преимущество UEBA-систем — способность обнаруживать ранние признаки готовящейся атаки, в то время как остальные системы пытаются предотвратить непосредственно саму атаку и поэтому "не имеют права на ошибку". Изучим рынок коммерческих UEBA-систем подробнее. Любые UEBA-системы описараются на три столба[3]:

- **Случаи использования**. Разработчики должны предусмотреть возможные инциденты, например, скомпроментированного пользователя, предателя.

- **Используемые данные**. Система может анализировать такие данные, как логи, сетевой трафик, используемый пользователями контент

- **Используемые методы аналитики**. По исследованию[3] на данный момент UEBA-системы имеют следующие средства аналитики:

  – основаны на правилах

  – статистическое моделирование

  – классические алгоритмы машинного обучения

  По мнению аналитиков Gartner применение глубокого обучения только набирает обороты в данной области, а в будущем будут применяться ансамбли нейронных сетей и генеративно-состязательные сети.

Согласно отчету Gartner[3] самостоятельные UEBA-системы становятся менее востребованными, им на замену приходят современные SIEM-системы со встроенной UEBA.

Потенциально доступную для анализа поведенческую информацию можно разделить на два вида:

- Контекст — структурированные данные, например метки времени, когда пользователь открывал файл, адресат, которому он отправил письмо

- Контент —- нектруктурированные данные, с коротыми работает пользователь, например, содержимое писем пользователя

Gartner в [4] составили “магический квадрат” (рис. 1) компаний-производителей систем управления безопасностью и событиями. По горизонтальной оси отмечается “полнота видения” — способность комании удовлетворить широкие запросы потребителей, используя новейшие технологии, по вертикальной оси — “способность реализации” — качество предлагаемого продукта.

Рис. 1: Магический квадрат[4]

Рассмотрим возможности анализа поведения пользователя продуктов лидеров IBM и Splunk.

IBM QRadar SIEM

IBM QRadar SIEM - комплексное решение, вокруг которого построена платформа QRadar Security Intelligence Platform, включающая в себя множество компонентов. IBM QRadar User Behavior Analytics (UBA)[5] — дополнительный модуль в рамках данной платфомы. IBM QRadar UBA использует готоые правила и модели машинного обучения для анализа пользовательского контекста[5] . В техническом отчете[6] говорится об использовании лишь контекста пользователей и упоминаются использование контента. Также, говорится об использовании “множества” методов машинного обучения, но не раскрывается, какие именно. Использование нейронных сетей не упоминется, а значит, скорее всего они не используются. Среди интересных особенностей можно выделить:

- Возможность, как поверхностного анализа всех пользователей, так и глубокое наблюдение за отдельными подозрительными пользователями
- Создание и сравнение пользователей в разных списках наблюдения
- Полная интеграция с системой IBM QRadar SIEM
- Аггрегация данных журналов брандмауэра, информации о пользовании облачными приложениями, сетевых потоках и прочее.

Splunk UBA

Компания Splunk имеет самостоятельный продукт Splunk UBA для анализа поведения пользователей. В white paper [7] подход Splunk UBA четырехэтапный:

1. Сбор данных. Платформа Splunk —The Data-to-Everything Platform позволяет собирать огромное множество данных, включая такие данные как записи звонков, положение GPS и пр. О сборе контентных данных ничего не говорится.
2. Поиск аномалий. На этом шаге собираются аномальные факты поведения пользователей. Splunk используют обучение без учителя для детектирования аномалий(не уточняется, какие), причем модель обучается для каждого пользователя своя.
3. Аггрегация. На этом шаге анализируются большие объемы аномалий, найденных на втором шаге. Для этого используются предобученные алгоритмы машинного обучения без учителя (не уточняется, какие)
4. Исследование. Полученные результаты предоставляются аналитику в человекочитаемом виде для подробного анализа.

Из обзора коммерческих продуктов видно, что на данных момент даже лидеры рынка UEBA-систем не анализируют контент пользователей, а также используют только классические методы машиннного обучения для обнаружения аномалий. Контент пользователей представляется огромным массивом данных, который мог бы быть учтен при анализе поведения пользователя, а более современных подходы в машинном обучении могли бы улучшить качество определения угроз. Таким образом, исследование и разработка современных методов машиного обучения для обнаружения внутренних угроз, использующих как контекстных поведенческих данных, так и контентных, является актуальной задачей.

1 Постановка задачи

Формальная постановка задачи звучит следующим образом:

1. Исследование и разработка методов обнаружения аномального поведения пользователей с использованием методов машинного обучения на основе поведенческой информации следующих типов:
   (a) Контекст пользовательских операций
   (b) Контент файлов

2 Обзор существующих решений рассматриваемой задачи или ее модификаций

Рассмотрим существующие подходы к задаче обнаружения инсайдерских угроз, сделаем их сравнение. Как уже было описано ранее, поведенческие данные можно разделить на два вида: контент — неструктурированные данные и контекст — структурированные. В литуратуре можно встретить подходы использующие только контент([8], [9]), только контекст([10], [11], [12], [13]) и использующие оба вида([14], [15]) поведенческих данных.

В первом разделе рассмотрим предлагаемые в литературе подходы к решению задачи обнаружения инсайдеров. Затем обратим особое внимание на способы использования в работах контентной составляющей поведенческих данных, сравним используемые в различных работах подходы. В конце, рассмотрим существующие для данной работы наборы данных.

2.1 Обзор существующих подходов

Базовый подход к любой задаче обнаружения чего-либо - придумать правила. Так поступили авторы [8]. В этой работе была предложена иерархия инсайдеров и определены присущие им черты характера. Затем по контенту пользовательского поведения с помощью сервиса IBM Personality Insights были определены черты характера пользователей, которые классифицировались по предложенным правилам.

Задача обнаружения инсайдерских угроз может быть рассмотрена с двух сторон: как задача обнаружения аномалий, и как задача обучения с учителем. Оба имеют недостатки:

- Не каждое аномальное поведение представляет собой угрозу.
- Для обучения алгоритма с учителем необходима разметка. На данный момент есть проблема с размеченными наборами данных.

Таккже заметим, что поведение пользователя представляет собою последовательность его действий, напоминающую текст. Поэтому во многих работах применяются хорошо показавшие себя на задачах обработки естественного языка (англ. NLP - Natural Language Processing) идеи: использование рекуррентных нейросетей, использование механизма внимания (англ. Attention).

2.1.1 Обнаружение инсайдеров как задача поиска аномалий

В [10] авторы для каждого сотрудника обучили LSTM-автокодировщик его “нормальному” поведению, а также построили “граф” сообщества пользователей, в кото-

ром ребра - общение с помощью писем . Затем они разделили всех пользователей на непересекающиеся сообщества Лувенским методом[16] и посчитали среднюю ошибку реконструкции для каждого пользователя на всех обученных маделях его группы. Чем больше эта ошибка, тем более аномально поведение пользователя.

В [14] авторы моделируют поведение каждого пользователя с трех сторон:

- Аггрегация дневной активности - контекст пользовательского поведения за каждый день
- LDA-моделирование контента электронных писем
- Положение пользователя в грае коммуникации внутри компании

Затем применяют четыре алгоритма детекции аномалий:

- Метод главных компонент. В качестве значения аномалий использовалась ошибка реконструкции.
- Метод K ближайших соседей.
- Оценка параметров нормального распределения на тренировочной выборке и последующая оценка вероятности того, что новые наблюдения принадлежат оцененному распределению.
- Оценка плотности распределения окном Парзена и последующая оценка вероятности принадлежности новых наблюдений данному.

В работе [17] для нахождения аномалий также использовались классические алгоритмы: Изолирующий лес и Однокласовый метод опорных векторов, агрегация производилась по нескольким промежуткам времени. Авторы попытались учесть последовательных характер поведения пользователей: строилось несколько моделей по периодам времени и каждая последующая модель получала особый *trust score* от предыдущий модели.

В работе [9] использовался только контент поведенческих данных: применяется аспектно-ориентированный анализ эмоциональной окраски (англ. ABSA — Aspect-based sentiment analysis) на контентных поведенческих данных с помощью собственной сложной рекурретной модели Для обнаружения аномалий используется Изолирующий лес.

В работе [18] авторы исследовали применение механизма внимания для RNN, GRU и LSTM архитектур для обнаружения инсайдерских угроз. Как показали эксперименты, наилучшее значение ROC AUC показало применение механизма внимания для RNN и LSTM сетей.

2.1.2 Обнаружение инсайдеров как задача обучения с учителем

В [15] авторы пробуют применить порядка сорока методов классификации и приходят к выводу, что лучше всего в этой задаче себя показывает случайный лес. В качестве признакового пространства в работе использовались эмоциональный факторы из писем и контекстные данные. Инетересно, что пользоватлеи классифицировались сразу на пять классов: добропорядочный, бывший работник, вор, тот, кто сливает информацию, и саботажник.

В [11] авторы пробуют применить сразу четыре алгоритма обучения с учителем: Логистическую регрессию, Случайный лес, Нейронную сеть(архитектура не показана) и XGBoost. В работе используются только контекстные поведенческие данный, которые агрегируются на разных масштабах: от недели до N действий пользователя в течение сессии. Авторы исследуют эффективность агрегации контекста пользователя на разных промежутках времени и приходят к выводу, что наиболее эффективно аггрегировать поведенческий контекст в течение сессии или рабочего дня.

В [12] предлагается двухэтапный подход: сначала авторы обучают LSTM-сеть поведению пользователей, а затем извлекают из нее признаки и подают их на вход сверточной сети-классификатору. В [19] авторы применяют обратный подход: сначала сверточной сетью извлекаются признаки, а затем они подаются на вход рекуррентному классификатору с LSTM-ячейками.

Еще один нейросетевой подход описывается в [13]. Для обнаружения инсайдеров предлагается применять GRU-классификатор.

Оригинальный подход предложен в [20]. В работе авторы сконструировали 20 признаков из контекстных поведенческих данных преложенным в [21] способе. Затем по вектору признаков составили изображение: значения признаков перевели в диапазон 0-255 и растянули в изображение 32x32. Для классификации доучивались предобученные VGG16, Inception и Mobilenet на полученных изображениях.

2.1.3 Подходы к работе с контекстом

One-hot, агрегация, из статьи [21]

2.1.4 Подходы к работе с контентом

Если контекст достаточно агрегировать и закодировать, то контент имеет сликом большой объем для простого кодирования я виде последовательности. Рассмотрим встречающиеся в литературе подходы к использованию контента.

В работе [9] использовался аспектно-ориентированный анализ эмоциональной окраски(ABSA)для получения представления контентных данных. В первом случае использовалась современная нейросетевая рекуррентная модель с механизмом внимания.

Таблица 1: Сравнение статей, рассматривающих задачу обнаружения инсайдеров, как задачу поиска аномалий

Статья	Контент	Контекст	Метод обнаружения	Наборы данных для тестирования	Качество
LAC[10]	-	+	Вычисление ошибки реконструкции LSTM AUTOENCODER	CERT v6.2	???
Behavior Modeling + Anomaly Detection [14]	+	+	KNN, PCA, статистических методов	CERT v6.2	•
ABSA model + IF [9]	+	-	Изолирующий лес	Enron, Enron+	•
Attention-based [18]	-	+	•	CERT v4.2	•
Trust aware unsupervised [17]	-	+	One Class SVM, Isolation Forest	CERT v4.2	•

В работе [15] использовался список AFINN-111 для определения настроений[22] содержимого писем и посещенных сайтов в течение месяца и формировался общий индекс риска для содержимого писем и сайтов соответственно.

В работе [14] авторы применяли LDA модель для тематического моделирования содержимого пользовательских писем. Каждое письмо представлется вектором из 50 тематик и входит в "контентную"модель пользовательского поведения, затем для этих векторов применялись описанные выше алгоримы обнаружения аномалий.

2.1.5 Сравнение работ

В таблицах 1 и 2 показано общее сравненние рассмотренных подходов. Сравнение приведенного в статьях качества сложно сделать из-за разных версий наборов данных, разных постановок экспериментов и различных метрик качества.

Таблица 2: Сравнение статей, рассматривающих задачу обнаружения инсайдеров, как задачу обучения с учителем

Статья	Контент	Контекст	Метод обнаружения	Наборы данных для тестирования	Качество
LSTM + CNN [12]	-	+	CNN-классификатор	CERT v4.2	AUC=0.9449
CNN + LSTM [19]	-	+	Классификатор скрытого состояния LSTM	CERT r4.2	•
Analyzing Data Granularity Levels for Insider Threat Detection using Machine Learning [11]	-	+	LogReg, Random Forest, NN и XGBoost	CERT r5.2 - тренировка, CERT r5.1, r6.2 - тест	•
GRU [13]	-	+	GRU-классификатор	CERT v4.2	точность - 0.92
Classifier Suites [15]	+	+	Порядка 40 классических классификаторов	CERT v4.2	•
Image-Based features [20]	-	+	CNN-классификатор	CERT v4.2	•

2.2 Обзор наборов данных

Список литературы

[1] A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations / N. Al-Mhiqan, Mohammed, Rabiah Ahmad et al. // Applied Sciences. — 2020. — 07. — Vol. 10. — P. 1–41.

[2] 2020 Cost of Insider Threats: Global Report : Rep. ; Executor: Ponemon Institute. — New York NY : 2020.

[3] Market Guide for User and Entity Behavior Analytics : Rep. ; Executor: Gorka Sadowski, Jonathan Care, Neil MacDonald, Henrique Teixeira. — New York NY : 2019. — 05.

[4] Magic Quadrant for Security Information and Event Management : Rep. ; Executor: Kelly Kavanagh, Toby Bussa, Gorka Sadowski. — New York NY : 2020. — 02.

[5] IBM QRadar User Behavior Analytics. — Access mode: `https://www.ibm.com/products/qradar-user-behavior-analytics`.

[6] IBM QRadar 7.3.1 with UBA 3.0 : Rep. ; Executor: Chris Kissel. — New York NY : 2019. — March.

[7] White paper: Splunk Integrated Behavior Analytics : Rep. : 2020.

[8] Eftimie S., Moinescu R., Răcuciu C. Insider Threat Detection Using Natural Language Processing and Personality Profiles // 2020 13th International Conference on Communications (COMM). — 2020. — P. 325–330.

[9] Employee profiling via aspect-based sentiment and network for insider threats detection // Expert Systems with Applications. — 2019. — Vol. 135. — P. 351 – 361. — Access mode: `http://www.sciencedirect.com/science/article/pii/S0957417419303781`.

[10] Paul Sudipta, Mishra Subhankar. LAC : LSTM AUTOENCODER with Community for Insider Threat Detection. — 2020. — 2008.05646.

[11] Le D. C., Zincir-Heywood N., Heywood M. I. Analyzing Data Granularity Levels for Insider Threat Detection Using Machine Learning // IEEE Transactions on Network and Service Management. — 2020. — Vol. 17, no. 1. — P. 30–44.

[12] Insider Threat Detection with Deep Neural Network / Fangfang Yuan, Yanan Cao, Yanmin Shang et al. // Computational Science – ICCS 2018 / Ed. by Yong Shi, Haohuan Fu, Yingjie Tian et al. — Cham : Springer International Publishing, 2018. — P. 43–54.

[13] New insider threat detection method based on recurrent neural networks / N. Al-Mhiqani Mohammed, Ahmad Rabiah, Zainal Abidin Zaheera et al. // Indonesian Journal of Electrical Engineering and Computer Science. — 2020. — 03. — Vol. 17. — P. 1474–1479.

[14] Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms / Taewoo Park Kim, Nam Kyu, Cho Honghyun, Kang // Applied Sciences. — 2019. — 09. — Vol. 9. — P. 4018.

[15] Noever David. Classifier Suites for Insider Threat Detection. — 2019. — 1901.10948.

[16] Fast unfolding of communities in large networks / Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre // Journal of Statistical Mechanics: Theory and Experiment. — 2008. — Oct. — Vol. 2008, no. 10. — P. P10008. — Access mode: http://dx.doi.org/10.1088/1742-5468/2008/10/P10008.

[17] Aldairi Maryam, Karimi Leila, Joshi James. A Trust Aware Unsupervised Learning Approach for Insider Threat Detection. — 2019. — 07. — P. 89–98.

[18] Attention-Based LSTM for Insider Threat Detection / Yuan Fangfang, Shang Yanmin, Liu Yanbing et al. — 2019. — 11. — P. 192–201. — ISBN: 978-981-15-0870-7.

[19] Insider Threats Detection using CNN-LSTM Model / Ahmed Saaudi, Zaid Al-Ibadi, Yan Tong, Csilla Farkas. — 2019. — 04. — P. 7.

[20] G Gayathri R, Sajjanhar Atul, Xiang Yong. Image-Based Feature Representation for Insider Threat Classification. — 2019. — 1911.05879.

[21] Chattopadhyay P., Wang L., Tan Y. Scenario-Based Insider Threat Detection From Cyber Activities // IEEE Transactions on Computational Social Systems. — 2018. — Vol. 5, no. 3. — P. 660–675.

[22] Äruprup Nielsen Finn. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs // CoRR. — 2011. — Vol. abs/1103.2903. — 1103.2903.