УДК 004.021

Лебедев Дмитрий Валерьевич,
магистрант,
кафедра анализа систем и принятия решений,
Институт экономики и управления,
ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина,
г. Екатеринбург, Российская Федерация

ПРОГНОЗИРОВАНИЕ СЦЕНАРИЕВ КИБЕРАТАК С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ БЕЗ УЧИТЕЛЯ

Аннотация:
На сегодняшний день прогнозирование кибератак можно рассматривать с помощью использования технологий интеллектуального анализа данных (Data Minnig) и машинного обучения (Machine Learning). Уже сейчас имеются определенные способы прогнозирования действий злоумышленника. Однако данные варианты являются сложными для понимания и их интеграции в системы по информационной безопасности.
В данной статье представлен метод, позволяющий определить дальнейшие тактики и техники злоумышленника с целью своевременного реагирования и принятия защитных мероприятий. Метод представляет собой ассоциативный поиск правил корреляции последовательных действий злоумышленника.

Ключевые слова:
Data Mining, ассоциативные правила, алгоритм Apriori, алгоритм FP-Growth, способы реализации атак, MITRE ATT&CK.

Введение:

В соответствии с рядом различных стандартов по обеспечению информационной безопасности (далее ИБ), для противодействия угрозам ИБ, снижения рисков ИБ и управлению инцидентами ИБ необходимо обеспечивать достаточный для организации уровень ИБ на протяжении длительного времени.

Актуальность обеспечения ИБ заключается в тенденции роста угроз и инцидентов в информационном пространстве. Согласно отчету Positive Technologies за 2022 год общее количество инцидентов ИБ увеличилось на 20,8% по сравнению с 2021 годом. В 2020 году рост киберинцедентов составил 51%, а в 2021 – всего 6,5% [1, 2, 3]. На рисунке 1 представлен график по количеству инцидентов в 2019, 2020, 2021 и 2022 годах (по кварталам).

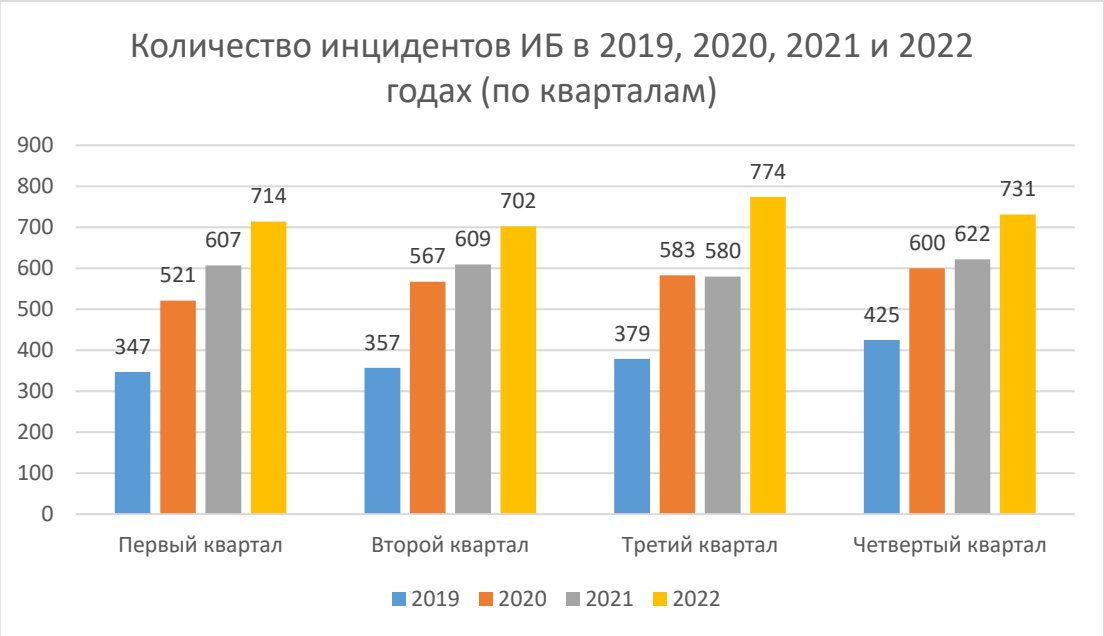

Рисунок 4 – Количество инцидентов ИБ в 2019, 2020, 2021 и 2022 годах по кварталам

Наблюдается тенденция по ежегодному увеличению инцидентов ИБ. Таким образом необходимо иметь инструменты по противодействию злоумышленникам в цифровом и информационном пространствах.

Важно учитывать, что ИБ на предприятии должна строиться вокруг стабильности существования бизнеса и для поддержания непрерывности бизнес-процессов. Но с учетом роста угроз ИБ и кибератак стоит уделять внимание возможности заранее спрогнозировать действия злоумышленников.

Целью данной работы являлась возможность спрогнозировать цепочку действий злоумышленника (сценарий кибератаки) с помощью использования инструментов машинного обучения без учителя – методов Data Mining.

Материалы и методы:

На основании анализа данных об известных сценариях кибератак можно спрогнозировать действия злоумышленника по его первого шагу.

Для определения данной информации необходимо получить данные (выборку) по известным сценариям кибератак. Как правило, такие сценарии представляют собой поэтапное описание действий злоумышленников с использованием тактик и техник MITRE ATT&CK. Пример такого сценария приведен в таблице 1, где действия — это техники злоумышленников по матрице MITRE ATT&CK.

Таблица 1 – Пример сценария кибератаки

Номер сценария	**Действие 1**	**…**	**Действие x**
1	T1059.002	…	T1219
…	…	…	…
m	T1566.001	…	T1578.003

Данные сценарии были получены из открытых источников:

- профессиональные социальные сети;
- открытые отчеты по инцидентам ИБ.

Для возможности спрогнозировать сценарий кибератаки использовались методы машинного обучения без учителя – методы Data Mining – поиск ассоциативных правил. Таким образом данный метод прогнозирования сценария кибератаки основан на принципе «если …, то …».

Для поиска ассоциативных правил использовались два алгоритма:

1. Apriori;
2. FP-Growth.

Псевдокод алгоритма Apriori и FP-Growth приведен на рисунках 2 и 3 соответственно [4].

вход: X^ℓ — обучающая выборка; $\delta = \texttt{MinSupp}$; $\varkappa = \texttt{MinConf}$;
выход: $R = \{(\varphi, y)\}$ — список ассоциативных правил;

множество всех частых исходных признаков:
$G_1 := \{f \in \mathscr{F} \mid \nu(f) \geqslant \delta\}$;
для всех $j = 2, \ldots, n$
  множество всех частых наборов мощности j:
  $G_j := \{\varphi \cup \{f\} \mid \varphi \in G_{j-1},\ f \in G_1 \backslash \varphi,\ \nu(\varphi \cup \{f\}) \geqslant \delta\}$;
  если $G_j = \varnothing$ **то**
    выход из цикла по j;
$R := \varnothing$;
для всех $\psi \in G_j,\ j = 2, \ldots, n$
  AssocRules $(R, \psi, \varnothing)$;

функция AssocRules (R, φ, y)
  вход: (φ, y) — ассоциативное правило;
  выход: R — список ассоциативных правил;
  для всех $f \in \varphi$: $\mathrm{id}_f > \max\limits_{g \in y} \mathrm{id}_g$ (чтобы избежать повторов y)
    $\varphi' := \varphi \setminus \{f\}; \quad y' := y \cup \{f\}$;
    если $\nu(y'|\varphi') \geqslant \varkappa$ **то**
      добавить ассоциативное правило (φ', y') в список R;
      если $|\varphi'| > 1$ **то**
        AssocRules (R, φ', y');

Рисунок 5 – Псевдокод алгоритма Apriori

вход: X^ℓ — обучающая выборка;
выход: FP-дерево T; $\langle f_v, c_v, S_v \rangle$ для всех вершин $v \in T$;

упорядочить признаки $f \in \mathcal{F}$: $\nu(f) \geqslant \delta$ по убыванию $\nu(f)$;
ЭТАП 1: построение FP-дерева T по выборке X^ℓ
для всех $i := 1, \ldots, \ell$
  $v := v_0$;
  для всех $f \in \mathcal{F}$ таких, что $f(x_i) = 1$, по убыванию $\nu(f)$
    если нет дочерней вершины $u \in S_v$: $f_u = f$ **то**
      создать новую вершину u; $S_v := S_v \cup \{u\}$;
      $f_u := f$; $c_u := 0$; $S_u := \varnothing$;
    $c_u := c_u + 1$; $v := u$;
ЭТАП 2: рекурсивный поиск частых наборов по FP-дереву T
FP-find$(T, \varnothing, \varnothing)$;

функция FP-find (T, φ, R)
  вход: FP-дерево T, частый набор φ, список наборов R;
  выход: добавить в R все частые наборы, содержащие φ;
  для всех $f \in \mathcal{F}$: $V(T, f) \neq \varnothing$ по уровням снизу вверх
    если $C(T, f) \geqslant \ell\delta$ **то**
      добавить частый набор $\varphi \cup \{f\}$ в список R;
      $T' := T|f$ — условное FP-дерево;
      найти по T' все частые наборы, включающие φ и f:
      FP-find$(T', \varphi \cup \{f\}, R)$;

Рисунок 6 – Псевдокод алгоритма FP-growth

Сравнение эффективности использования данных алгоритмов в общей задаче поиска ассоциативных правил представлено на рисунке 4 [4].

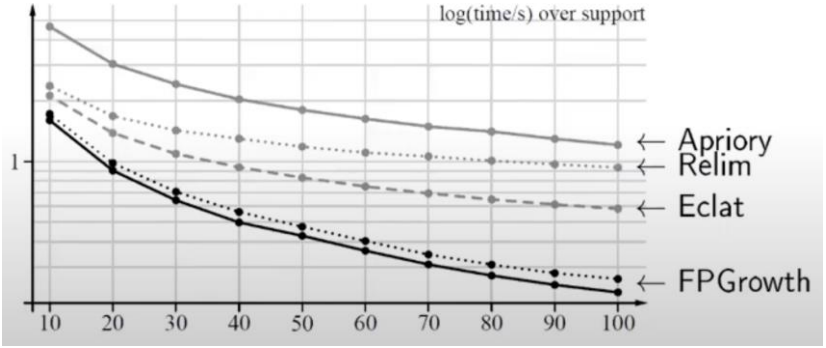

Рисунок 7 – Сравнение эффективности использования алгоритмов поиска ассоциативных правил

Использование данных алгоритмов позволило получить ассоциативные правила типа «если злоумышленник использовал технику 1, то дальнейшая вероятная его техника – 2».

Результаты:

По результатам использования алгоритмов Apriori и FP-Growth были сформированы таблицы, позволяющие определить дальнейшие действия злоумышленника по его первому шагу. Вырезка из таких таблиц приведена в таблице 2.

Таблица 2 – Результат выполнения алгоритмов

№	Left hand side	Right hand side	Support	Confidence	Lift
1	T1027	T1014	0,036231884	0,714285714	2,053571429
2	T1027	T1497.001	0,02173913	0,75	2,15625
3	T1027, T1014	T1059.001	0,02173913	1	2,875
…	…	…	…	…	…
n	T1036.005	T1140	0,028985507	1	3,209302326

В полученных ассоциативных правилах важными параметрами являются поддержка (Support), доверие (Confidence), подъем (Lift).

Support представляет собой базовую вероятность того, что событие произойдет. Она измеряется долей транзакций, в которых появляется набор элементов. Формула для определения support:

$$support(A \to B) = P(A \cup B)$$

Confidence определяет вероятность наступления события 1 при условии, что событие 2 уже произошло, т.е. доверие – это условная вероятность. Формула для определения confidence:

$$confidence(A \to B) = \frac{P(A \cup B)}{P(A)}$$

Lift представляет собой вероятность того, что событие 1 произойдет при условии того, что произойдет событие 2, при этом учитывается вероятность обоих событий. Lift показывает зависимость двух событий. Формула для определения lift:

$$lift(A \to B) = \frac{confidence(A \to B)}{support(B)}$$

В задаче, когда необходимо по первому действию злоумышленника определить дальнейшее развитие атаки рекомендуется использовать алгоритм Apriori. Однако, когда злоумышленник уже совершил определенную цепочку действий и необходимо дальнейшее прогнозирование, рекомендуется использовать алгоритм FP-Growth. Это связано с тем, что алгоритм FP-Growth является более эффективным для быстрого поиска частных наборов на больших данных.

Выводы:

Таким образом, в данной работе был продемонстрирован метод прогнозирования цепочки действий злоумышленников по его первому шагу с использованием методов машинного обучения без учителя. Данный метод позволяет заранее определить дальнейшие возможные действия злоумышленника для того, чтобы можно было своевременно принять защитные меры.

Однако, используя данный метод прогнозирования кибератак, нужно учитывать, что злоумышленники постоянно меняют свои техники для реализации одних и тех же атак. Поэтому, для использования данного метода рекомендуется использовать временные характеристики в обучающем датасете, например, анализировать атаки за последние полгода.

Дальнейшее использование данного алгоритма для прогнозирования кибератак требует подтверждения эффективности, например, через проведение испытаний на киберполигоне.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Positive Technologies. Исследования. Аналитика. Актуальные киберугрозы: итоги 2022 года. [Электронный ресурс]. – Режим доступа: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2022/, свободный. (дата обращения 04.10.2023)
2. Positive Technologies. Исследования. Аналитика. Актуальные киберугрозы: итоги 2021 года. [Электронный ресурс]. – Режим доступа: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2021/, свободный. (дата обращения 04.10.2023)
3. Positive Technologies. Исследования. Аналитика. Актуальные киберугрозы: итоги 2020 года. [Электронный ресурс]. – Режим доступа: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2020/, свободный. (дата обращения 04.10.2023)
4. Машинное обучение (курс лекций, К.В. Воронцов). [Электронный ресурс]. – Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_%28курс_лекций%2C_К.В.Воронцов%29, свободный. (дата обращения 14.04.2023)

Lebedev Dmitry Valerievich,
master's student,
department of systems analysis and decision-making,
Graduate School of Economics and Management,
Ural Federal University named after the First President of Russia B.N. Yeltsin,
Yekaterinburg, Russian Federation

FORECASTING CYBERATTACK SCENARIOS USING UNSUPERVISED MACHINE LEARNING METHODS

Abstract:
The forecasting of cyberattacks can be considered using data mining technologies and machine learning. There are already certain ways to predict the actions of an attacker. However, these options are difficult to understand and integrate into information security systems.
This article presents a method that allows you to determine further tactics and techniques of the attacker in order to respond in a timely manner and take protective measures. The method is a search for associative rules of correlating sequential actions of the attacker.

Keywords:
Data Mining, associative rules, Apriori algorithm, FP-Growth algorithm, methods of implementing attacks, MITRE ATT&CK.