КІБЕРБЕЗПЕКА ТА ЗАХИСТ КРИТИЧНОЇ ІНФОРМАЦІЙНОЇ ІНФРАСТРУКТУРИ / CYBERSECURITY & CRITICAL INFORMATION INFRASTRUCTURE PROTECTION

DOI: 10.18372/2225-5036.26.15569

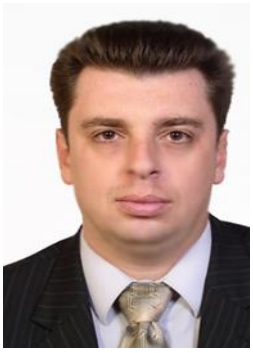
МЕТОДЫ РАСПОЗНАВАНИЯ КИБЕРАТАК С УЧЕТОМ МОНИТОРИНГА ИНФОРМАЦИОННОЙ СРЕДЫ

Хорошко В.А.[1], Браиловский Н.Н.[2]

[1]Национальный авиационный университет
[2]Киевский национальный университет имени Тараса Шевченка

ХОРОШКО Володимир Олексійович, д.т.н., професор.

Рік та місце народження: 1945 рік, м. Харків, Україна.
Освіта: Київський інститут інженерів цивільної авіації, 1968 рік.
Посада: професор кафедри безпеки інформаційних технологій.
Наукові інтереси: інформаційна безпека, технічні системи захисту інформації, аналіз функціонування складних систем.
Публікації: більше 500 наукових публікацій, серед яких наукові статті, монографії, підручники та навчально-методичні посібники.
E-mail: professor_va@ukr.net.
Orcid ID: 0000-0001-6213-7086.



БРАІЛОВСЬКИЙ Микола Миколайович, к.т.н., доцент

Рік народження: 1972, м. Київ, Україна.
Освіта: Українська державна академія зв'язку ім. О.С. Попова, 1994 рік.
Посада: доцент кафедри кібербезпеки та захисту інформації.
Наукові інтереси: національна безпека, методи та засоби технічного захисту інформації, захист кіберпростору, соціальна інженерія.
Публікації: понад 130 наукових публікацій, серед яких наукові статті, колективні монографії, тези та матеріали доповідей на конференціях, підручники та науково-методичні посібники.
E-mail: bk1972@ukr.net.
Orcid ID: 0000-0002-3148-1148.

Аннотация. *На сегодняшний день выявление кибератак является весьма актуальной задачей. Для этой цели используется мониторинг сетей. Причем, при этом возникает необходимость оперативной аналитической обработки информации, требующая применения методов интеллектуального анализа данных. Интеллектуальный анализ данных помогает извлечению знаний из полученных данных. Цель применения интеллектуального анализа данных к решению задач мониторинга кибератак – получение ранее не известных, нетривиальных, доступных для интерпретации процессов знаний, закономерностей в мониторинге, то есть – данных, полезных для поддержания принятия решений. Неотъемлемой частью распознающей системы является обучение, имеющее конечной целью формирование эталонных описаний классов, форма которых определяется способом их использования в решающих правилах, а также выбор информационных признаков для распознавания этих эталонных классов. При написании данной работы сделана попытка изложить в определенной логической последовательности основные аналитические методы распознавания кибератак в современных условиях кибервойны с учетом мониторинга информационной среды. Приведён перечень факторов, подтверждающих целесообразность применения методов распознавания образов для анализа данных мониторинга атак. Кроме того, рассмотрены меры сходства, которые используются в алгоритмах ранжирования и кластеризации кибератак. Показано, что целесообразность их применения зависит от конкретных задач.*

Ключевые слова: *кибератака, мониторинг сетей, методы интеллектуального анализа, методы распознавания образов, системы киберзащиты.*

Вступление

Постоянное возрастание роли информационной сферы на современном этапе характеризует развитие общества. По структуре оно представляет собой совокупность информации, информационных систем и информационных связей объектов, которые производят подготовку информации, ее хранение, распространение и использование, а также процессов регулирования возникающих конфликтов в этих общественных отношениях.

Информационная сфера стала сегодня базой для развития всех других сфер в жизни человека, общества и государства.

В информационной сфере происходят различные события и явления, анализ которых становится жизненно необходимым для любого объекта.

В современных условиях все больше распространяется аксиома, что киберзащита информационных технологий должна по своим характеристикам соответствовать масштабам угроз и рисков. Отклонение от этого правила приведет к дополнительным убыткам. Для каждой информационной системы должен существовать оптимальный уровень киберзащищенности, который необходимо постоянно поддерживать.

Нет сомнений, киберзащита очень важна для информационных систем. Однако нет ответа на очень важный вопрос – насколько решения, которые предлагаются и/или реализуются, действительно соответствуют требованиям киберзащиты.

Всем специалистам в области киберзащиты известны основные постулаты, которые весьма актуальны и на сегодняшний день [1]:

– абсолютной киберзащиты информации создать невозможно;

– система киберзащиты информации должна быть комплексной;

– система киберзащиты информации должна быть адекватной к изменениям обстановки;

– система киберзащиты информации должна быть системою, а не просто набором хаотичных средств;

– системный подход к киберзащите информации должен применяться, начиная с этапа подготовки технического задания и заканчиваться оценкой эффективности и качества системы в процессе ее эксплуатации.

Следует учитывать, что система киберзащиты информации должна иметь целевое назначение. Причем, чем более конкретно сформулирована цель киберзащиты информации, детально выяснены ресурсы, которые имеются и определен комплекс ограничений, тем в большей степени возможно получить позитивный результат.

Однако следует отметить, что основным элементом внешнего воздействия на информационную систему является кибератака.

В последние годы кибератаки (КА) стали широко применяться не только отдельными хакерами или объединенными их в группы, но и государствами структурами некоторых стран.

Так, например, в 2008 году были осуществлены кибератаки Российской федерацией на банковскую систему Эстонии и государственные сайты Грузии, а с 2013 года осуществляются постоянные кибератаки на украинские сайты [2]. В настоящее время одним из самых сложных и полнофункциональных средств проведения атак на информационные системы являются кибератаки, которым присущи следующие функции [2]:

– распространение в сетях;

– перехват сетевых пакетов;

– обнаружение сетевых ресурсов и сбор перечня уязвимых паролей;

– передача информации на серверы злоумышленников;

– сканирование диска информационной системы на наличии определенных расписаний и контента;

– использование большого количества домен для приема команд с серверов управления.

Как сообщает газета The Financial Times, специалисты Центра правительственной связи (GCHQ) разведслужбы, отвечающие за безопасность британских властей, пришли к выводу, что ФСБ России используют программное обеспечение «Лаборатория Касперского» для скрытого наблюдения за сотрудниками правительства и военными Великобритании [1,2].

Кроме того, глава британского Национального центра безопасности (NCSC) заявил, что в 2016 году российские хакеры осуществили кибератаки на СМИ, систему телекоммуникаций и энергетический сектор Великобритании.

И, как сообщает английская газета The Daily Telegraph, глава NCSC отметил, что его ведомство взаимодействует с международными партнерами, представителями индустрии и общественными организациями для решения этой проблемы [1,2].

По данным NCSC с 2016 года его специалистам удалось предотвратить десятки миллионов кибератак. Они также приняли контрмеры в ответ на 590 хакерских атак, в том числе на масштабную кибератаку, происшедшую в мае 2017 года посредством вируса-вымогателя WannaCry, целью которого стала национальная система здравоохранения Великобритании. В этом нападении раньше подозревали хакеров, связанных с КНДР, однако, как пишет издание, имела место угроза, исходящая именно со стороны России [2].

Поэтому, **выявление кибератак является весьма актуальной задачей.** Для этой цели используется мониторинг. Причем, при этом возникает необходимость оперативной аналитической обработки информации, которая требует применения методов интеллектуального анализа данных.

В настоящее время ввод и хранение больших массивов данных мониторинга не представляют актуальных проблем.

На первое место выдвигаются ряд других вопросов: «Поможет ли эта информация выявлению кибератак? Как использовать «историю» мониторинговых данных, чтобы выявить кибератаки? Можно ли предсказать поведение таких процессов? и т.д.».

Эти вопросы становятся особенно важными при наличии больших массивов разнородных данных, какими являются данные мониторинга КА. По этой причине, интеллектуальный анализ данных помогает извлечению знаний из полученных данных. Цель применения интеллектуального анализа данных к решению

задач мониторинга КА – получение ранее не известных, нетривиальных, доступных для интерпретации процессов знаний, закономерностей в мониторинге, то есть – данных, полезных для поддержания принятия решений.

В литературе приводится классификация задач интеллектуального анализа данных по типам производимой информации [3,4] выделяя при этом пять основных видов задач.

1.Классификация (распознавание с «учителем») – наиболее распространённая задача интеллектуального анализа данных. Она позволяет выявить признаки, характеризующие однотипные группы КА – классы, для того, чтобы по известным значениям их характеристик можно было отнести к тому или иному классу. Ключевым моментом выполнения этой задачи является анализ множества классифицированных атак. В качестве методов решения задачи классификации могут использоваться различные алгоритмы [5,6], байсовские сети [7,8], индукция деревьев решений, индукция символьных правил [9,10], нейронные сети [11], параметрические алгоритмы распознавания [12] и другие.

2.Кластеризация. Результатом кластеризации является определение присущего исследуемым данным разбиения на кластеры. В большинстве случаев кластеризация субъективна, т.к. любой вариант разбиения на кластеры напрямую зависит от выбранной меры расстояния между кластеризуемыми атаками [3,13,14].

3.Выявление ассоциаций. Ассоциация определяется на основе свойств двух или нескольких одновременно наступающих событий. При этом производные правила указывают на то, что при наступлении одного события с той или иной степенью вероятности наступит и другое. Количественно сила ассоциации определяется несколькими величинами: предсказуемость, распространяемость и ожидаемая предсказуемость [15].

4.Выявление последовательностей. Подобно ассоциациям, последовательность имеет место между событиями, но наступающими не одновременно, а с некоторым определенным интервалом во времени. Таким образом, ассоциация является частным случаем последовательности с нулевым временным шагом [15,16].

5.Прогнозирование – на основе особенностей поведения текущих и предыдущих данных оцениваются будущие значения определенных числовых показателей. В задачах подобного типа наиболее часто используются традиционные методы математической аналитики, а также нейронные сети [17-21].

Целью данной работы является анализ и изложение в определенной логической последовательности основных аналитических методов распознавания кибератак в современных условиях кибервойны с учетом мониторинга информационной среды.

Основная часть

Пусть $X = \{x_1, x_2, \dots, x_n\}$ – начальное множество КА. Каждая атака, которая может быть осуществлена на информационную систему, описывается набором характеристик свойств-признаков $X' = (x_{i1}, x_{i2}, \dots, x_{im})$, i= *1,2,…,n*. Геометрически атаку удобно интерпретировать точкой или вектором в соответствующем m-мерном пространстве (информационной сфере). Строго говоря, такое описание атаки представляет собой ее образ.

Распознаванию подлежат не собственно атаки, а образы – формализованные понятия, с которыми ассоциируется КА. В дальнейшем для краткости изложения под термином «кибератака» будем понимать образ атаки.

Наличие определенной общности свойств у разных типов атак позволяет группировать атаки в некоторые подмножества $K_1, K_2, \dots, K_k$ множества X-классы.

Распознаваемые образы можно определить как отношения исходных атак $\{x_1, x_2, \dots, x_n\}$ к определенному классу K_j, j=*1,2,…,k*, с помощью выделения существенных признаков или свойств, характеризующих эти атаки. Это означает, что нужно построить однозначное отображение множества X на множестве классов $K = \{K_1, K_2, \dots, K_k\}; X \to K$.

Все k исследуемых классов представлены непересекающимися множествами своих «представителей»:

$$\{X_1^{(j)}, X_2^{(j)}, \dots, X_{n_k}^{(j)}\} \subset K_j; j = 1,2, \dots, k;\ n_1 + n_2 + \dots + n_k = n_0,$$

так называемыми обучающими выборками. Насколько хорошо атаки обучающих выборок отражают постоянную структуру классов, дает понятие представительности выборки.

Формально представительность обычно оценивается отношением n_j/m, где n_j – число атак в выборке, а m – размерность признакового пространства.

Используя информацию, содержащуюся в обучающейся выборке и, может быть, некоторые априорные данные о решаемой задаче, требуется построить решающее правило, наилучшим образом классифицирующее атаки распознавания.

Обычно решающее правило определяется разделяющей функцией R(x), которая в случае безошибочного разделения двух классов ведет себя следующим образом:

R(x)>0 для x ⊂ K_1,

R(x)<0 для x ⊂ K_2.

К априорным данным о решаемой задаче относятся:

- желаемая размерность;
- вероятностная ситуация;
- детерминированная ситуация;
- контрольная выборка.

Желаемая размерность. Под этим понятием подразумевается, что известно число признаков, необходимое для решения задачи. Во многих методах число признаков не задается, а определяется в процессе выбора пути оптимизации какого-либо критерия.

Вероятностная ситуация (задан вид распределения). Считается, что атака относится к данному классу с некоторой вероятностью. Такая ситуация порождает статистические методы.

Детерминированная ситуация (вид и параметры распределения не известны). Считается, что классы не пересекаются, т.е., КА различных классов изолированы друг от друга. Такая ситуация порождает непараметрические методы.

Контрольная выборка. Состоит из атак заданных классов:

$$\{X_1^{(j)}, X_2^{(j)}, \dots, X_{n_e}^{(j)}\} \subset K_j; j = 1,2, \dots, k;\ n_1 + n_2 + \dots + n_e = n_q.$$

Контрольные атаки не учитывают в процессе обучения, а используют для оценки качества распознавания. Объекты обучающей и контрольной выробок называют эталонными.

Если затраты, связанные с потерями от неправильного распознавания и с реализацией некоторого решающего правила обозначить С, то четверка множеств {X,K,R,C} будет характеризовать задачу распознавания. Можно выделить три типа задач:

1) задано множество классов K (обычно задается обучающая выборка), пространство признаков X, требуется найти решающее правило R, минимизирующее затраты С. Это задача распознавания при наличии обучения. Простым случаем этой задачи является ранжирование (случай k=1) – упорядочение видов атак по некоторой мере сходства относительно заданного класса;

2) задано множество классов K, решающие правило R, требуется найти систему признаков X, минимизирующих затраты С. Это задача минимизации пространства признаков;

3) задано пространство признаков X, требуется найти множество классов K и решающее правило R. Это задача кластеризации – разбиение атаки на некоторые, в общем случае не заданное, число классов в соответствии со свойствами самих атак.

Однако четверка множеств {X,K,R,C} характеризует тип задачи распознавания формально, так как не учитывает связи между входящими в нее величинами. Поэтому практически все методы распознавания основаны на определенных предположениях. Наиболее распространенными являются следующие [22]:

- о независимости выбора КА;
- о компактности классов;
- о существовании функции признаков каждого класса;
- о линейной разделимости классов;
- о нормальной распределенных матриц классов;
- о независимости признаков.

Использование гипотез такого рода предопределяет выбор алгоритма и качество решения задачи распознавания.

Важным понятием является отказ от распознавания. В сомнительных случаях (атака расположена слишком близко к разделяющей функции) или при нарушении гипотезы компактности (атака расположена далеко от средних значений классов) классификация не производится.

Используя отказ, можно повышать качество распознавания, хотя число правильных решений и уменьшается.

Понятие меры сходства позволяет оценить степень сходства между элементами распознающей системы, т.е. между атаками, классами атак, классом и отдельной КА.

В зависимости от решаемой задачи используется определенная мера сходства.

Неотъемлемой частью распознающей системы является обучение, имеющее конечной целью формирование эталонных описаний классов, форма которых определяется способом их использования в решающих правилах, а также выбор информационных признаков для распознавания этих эталонных классов.

Теперь рассмотрим типы исходных данных мониторинга.

Обозначим через X_i - КА, а через V_j - признаки (свойства, атрибуты) атак $V_j = X_i = (x_1, x_2, \dots, x_n)$ Можно выделить 4 типа исходных данных:

1. Таблица типа «атаки-признаки» (см. рис. 1) - для *n* атак заданы значения *m* признаков в фиксированный момент времени *t*.

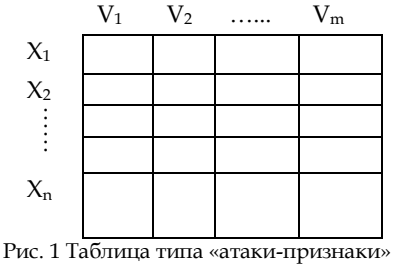


Рис. 1 Таблица типа «атаки-признаки»

2. Таблица значений одного признака (см. рис.2), измеренного в разные моменты времени $t_1, t_2, \dots, t_k$ для *n* атак.

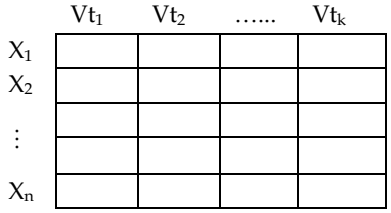


Рис. 2 Таблица значений одного признака

3. Таблица значений признаков одной атаки (см. рис. 3), измеренных в разные моменты времени $t_1, t_2, \dots, t_q$.

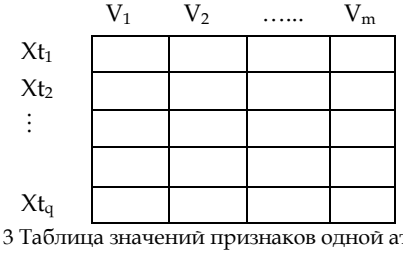


Рис. 3 Таблица значений признаков одной атаки

4. «Куб данных» - значения признаков (см. рис.4), измеренных в разные моменты времени $t_1, t_2, \dots, t_q$ для *n* атак.

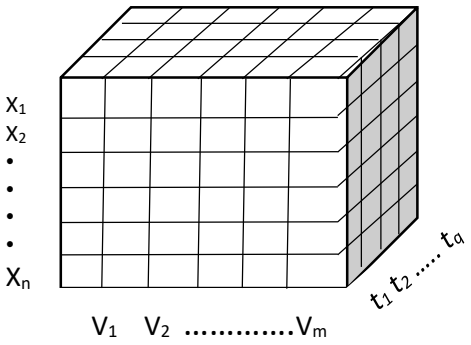


Рис. 4 «Куб данных»

Данные мониторинга характеризуются тем, что могут быть количественными и качественными, иметь разные размерности, пропущенные значения и погрешности измерения, носить сезонный характер, временной характер и т.п. Кроме того, часто переменные имеют разный диапазон измерений, так как измерены они разными методами или просто из-за того, что характеризуют разные свойства КА.

Непосредственное использование переменных в анализе может привести к тому, что классификацию будут определять те из них, которые имеют наибольший разброс значений. Поэтому применяются следующие виды стандартизации [12]:

1) «Z-шкалы». Из значений переменных вычитают их среднее, и эти значения делятся на стандартное отклонение:

$$x_{ij}^* = \frac{x_{ij}-\overline{x_j}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\overline{x_j})^2}},\ i=1,2,\ldots,n;\ j=1,2,\ldots,m.$$

2) «Разброс от 0 до 1». Линейным преобразованием переменных добиваются разброса значений от 0 до 1:

$$x_{ij}^* = \frac{x_{ij}-\min\limits_i x_{ij}}{\max\limits_i x_{ij}-\min\limits_i x_{ij}}.$$

3) «Максимум-1». Значения переменных делятся на их максимум:

$$x_{ij}^* = \frac{x_{ij}}{\max\limits_i x_{ij}}.$$

4) «Средние-1»: Значения переменных делятся на их среднее:

$$x_{ij}^* = \frac{x_{ij}}{\overline{x_{ij}}}.$$

5) «Стандартное отклонение – 1». Значение переменных делится на стандартное отклонение:

$$x_{ij}^* = \frac{x_{ij}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\overline{x_j})^2}}.$$

6) «Разброс от -1 до 1». Линейным преобразованием переменных добиваются разброса значений от -1 до 1:

$$x_{ij}^* = \frac{2x_{ij}-\min\limits_i x_{ij}-\max\limits_i x_{ij}}{\max\limits_i x_{ij}-\min\limits_i x_{ij}}.$$

7) Центрирование и масштабирование на среднее значение:

$$x_{ij}^* = \frac{x_{ij}-\overline{x_j}}{\overline{x_j}}.$$

Теперь рассмотрим меры сходства [23]

Причем, меры сходства двух элементов Z_i и Z_j определяются некоторой функцией $d(Z_i, Z_j)$, обладающей следующими свойствами:

1. Симметрия

$$d(Z_i, Z_j) = d(Z_j, Z_i);$$

2. Максимальное сходство элемента с самим собой

$$d(Z_i, Z_j) = 0, d(Z_i, Z_j) > 0, i \neq j.$$

Мера, удовлетворяющая этим свойствам, называется расстоянием, если выполняется неравенство треугольника

$$d(Z_i, Z_j) < d(Z_i, Z) + d(Z, Z_j).$$

Приведем наиболее распространенные меры сходства [3,12,14,23].

Узловым моментом в кластерном анализе считается выбор меры сходства атак, от которой решающим образом зависит окончательный вариант разбиения атак на группы при заданном алгоритме разбиения.

В каждом конкретной задаче этот выбор производится по-своему, с учетом главных целей исследования. Основные меры сходства между КА:

1) евклидово расстояние

$$d_E(X_i, X_j) = \left[\sum_{i=1}^{m}(x_{i1}-x_{j1})^2\right]^{1/2}; \quad (1)$$

2) взвешенное евклидово расстояние

$$d_B(X_i, X_j) = \left[\sum_{i=1}^{m}\omega_1(x_{i1}-x_{j1})^2\right]^{1/2}, \quad (2)$$

где ω_1 - весовой коэффициент;

3) потенциальная функция

$$d_\Pi(X_i, X_j) = \left[1+\propto d_E^2(X_i, X_j)\right]^{-1};\ \propto > 0 \quad (3)$$

$$\text{или } d'_\Pi(X_i, X_j) = \exp\left[-\propto d_E^2(X_i, X_j)\right] \quad (4)$$

$$\text{или } d''_\Pi(X_i, X_j) = \left|\frac{\sin d_E^2(X_i, X_j)}{\propto d_E^2(X_i, X_j)}\right|; \quad (5)$$

4) угол между векторами X_i и X_j

$$d(X_i, X_j) = arccos\frac{X_i * X_j}{|X_i| * |X_j|}. \quad (6)$$

Теперь рассмотрим меры сходства между классами.

1) Максимальное из исходных расстояний между КА разных классов

$$d_{min}(K_i, K_j) = \min d(X_1, X_m);\ X_1 \in K_i, X_m \in K_j\ . \quad (7)$$

2) Максимальное из исходных расстояний между КА равных классов

$$d_{max}(K_i, K_j) = \max d(X_1, X_m);\ X_1 \in K_i, X_m \in K_j\ . \quad (8)$$

3) Среднее значение парных расстояний между атаками разных классов

$$d_{mean}(K_i, K_j) = \frac{1}{n_i \cdot n_j}\sum_{X_1 \in K_i}\sum_{X_m \in K_j} d(X_1, X_m). \quad (9)$$

4) Расстояние, измеряемое по «центрам тяжести» классов

$$d_c(K_i, K_j) = d_E(\mu_i, \mu_j), \quad (10)$$

где $\mu_1 = \frac{1}{n_1}\sum_{X_j \in K_1} X_j$– вектор средних классов K_1.

5) Потенциальная функция

$$d_\Pi(K_i, K_j) = \frac{1}{n_i \cdot n_j}\sum_{X_1 \in K_i}\sum_{X_m \in K_j} d_\Pi(X_1, X_m). \quad (11)$$

6) Обобщенное (по Колмогорову) расстояние

$$d_K(K_i, K_j) = \left[\frac{1}{n_i \cdot n_j}\sum_{X_1 \in K_i}\sum_{X_m \in K_j} d_E^q(X_1, X_m)\right]^{1/q}. \quad (12)$$

Расстояния $d_{min}(K_i, K_j)$, $d_{max}(K_i, K_j)$, $d_{mean}(K_i, K_j)$ являются частными случаями обобщенного расстояния:

при $q \to \infty$ $d_K(K_i, K_j) = d_{max}(K_i, K_j)$,

при $q \to -\infty$ $d_K(K_i, K_j) = d_{min}(K_i, K_j)$,

при $q \to 1$ $d_K(K_i, K_j) = d_{mean}(K_i, K_j)$.

Рассмотрим теперь меры сходства между атаками и классами [23]:

1) Расстояние Махаланобиса

$$d_M(X, K_i) = (X-\mu_i)^T C_i^{-1}(X-\mu_i), \quad (13)$$

где μ_i и Ci – соответственно вектор средних и ковариационная матрица класса K_i;

2) Функция меры близости

$$d_{\text{ФМБ}}(X, K_i) = \left[\prod_{X_j \in K_i} d(X, X_j)\right]^{1/n_i}, \quad (14)$$

$$d_{\text{ФМБ}}(X, K_i) = \frac{1}{n_i}\sum_{X_j = K_i} \ln d(X, X_j),$$

где $d(X, X_j)$ – мера сходства между атаками, например, евклидово расстояние;

3) Потенциальная функция

$$d_\Pi(X, K_j) = \frac{1}{n_i}\sum_{X_j \in K_i} d_\Pi(X, X_j)\ , \quad (15)$$

где $d_\Pi(X_1, X_m)$ –определяется выражением (3);

4) Угловая мера подобия

$$d_{PSI}(X, K_i) = \left[\prod_{X_j \in K_i} sin(X \wedge X_j)\right]^{1/n_i}, \quad (16)$$

где $(X \wedge X_j)$ –угол между векторами X и X_j;

5) Расстояние до «центра тяжести» класса

$$d_c(X, K_i) = d_E(X, \mu_i), \quad (17)$$

где μ_i - вектор средних класса K_i;

6) Проекция на подпространство

$$d_{LS}(X, \pi_i) = \frac{|X_0|}{X}, \quad (18)$$

где $X_0 = \sum_{X_j \in K_i} \alpha_i X_i$ - проекция на гиперплоскость Г, натянутую на I ($I<n_i$) линейно независимых атак класса K_i,

$I<m$, m – размерность пространства,

α_j – коэффициенты, находятся из условия ортогональности (X-X_0, X_j)=0, j=1,2,…,I, $X_j \in K_i$.

Мера различия, соответствующая $d_{LS}(X, K_i)$

$$d_{LS}(X, K_i) = |X - X_0| \Big/ |X| = [|X|^2 + |X_0|^2 - 2(X, X_0)]^{1/2} \Big/ |X|. \quad (19)$$

Выводы

Целесообразность применения методов распознавания образов для анализа данных мониторинга КА обусловлена следующими факторами:

1. Одни и те же методы распознавания образов можно использовать для широкого класса различных по содержанию задач.

2. Класс используемых в анализе данных мониторинга методов распознавания образов обширен. Это методы ранжирования КА (признаков КА), методов распознавания «с учителем», методы кластеризации, методы группового учета аргументов. Каждой группе методов распознавания соответствует определенные задачи мониторинга. Но при этом в каждом конкретном случае применяются методы, учитывающие особенности исходных данных.

3. Применение методов распознавания образов позволяет проводить комплексный анализ разнородных данных, которые получают в процессе мониторинга.

Кроме того, рассмотренные меры сходства используются в алгоритмах ранжирования и кластеризации КА. Целесообразность их применения зависит от конкретной задачи.

В частности, евклидовое расстояние лучше использовать для количественных переменных, расстояние хи-квадрат – для исследования частотных таблиц, а также имеется множество мер для бинарных переменных.

Литература

[1] Грищук Р.В. *Основи кібернетичної безпеки*/ Р.В. Грищук, Ю.Г. Даник – Житомир: ЖНАЕУ, 2016. – 616 с.

[2] Пирцхалава Л.Г. *Информационное противоборство в современных условиях*/ Л.Г. Парцхалава, В.А. Хорошко, Ю.Е. Хохлачева, М.Е. Шелест – К: ЦП «Компринт», 2019- 226 с.

[3] Дюк В. *Data Mining* / В.Дюк, А. Сомойленко – СПб: Питер, 2001 – 368 с.

[4] Fuernkranz J. A *Brief Introduction to Knowledge Discovery in Databases* // OEGAI Journal. - 2005. - № 14(4). – pp. 14-17.

[5] Ту Дж *Принципы распознавания образов*. Изд. 2-е / Дж Ту, Р. Гонсалес – М.: Мир, 2001. – 412 с.

[6] Фукунага К. *Введение в статистическую теорию распознавания образов*. Изд. 3-е допол. / К. Фукунага – М.: Наука, 2005. – 388 с.

[7] Тулупьев А.Л. *Алгебраические байесовские сети. Логико-вероятностный подход к моделированию баз знаний с неопределенностью* / А.Л. Тулупьев – СПб.: СПИИРАН, 2000. – 292 с.

[8] Friedman N., Geiger D., Goldszmidt M., etc. *Bayesian Network Classifiers*//Machine Learning.-2007.-pp. 131-165.

[9] Michalski R. S. *A theory and methodology of inductive learning*//Artificial Intelligence.-1999.-20(2). - pp.111-162.

[10] Quinlan J. R. *Induction of decision trees* // Machine Learning. - 2006. - №1. - pp. 81-106.

[11] Fausett L. V. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. - Englewood Cliffs, New Jersey: Prentice Hall, 1994. - 461 p.

[12] *Классификация многомерных наблюдений Изд. 2-е* / С.А. Айвазян, З.И. Бежаева, О.В. Староверов. - Москва: Статистика, 2001. – 244 с.

[13] Дюран Б., Оделл П. *Кластерный анализ*. Изд. 3-е доп./Б. Дюран, П. Оделл.-М.: Статистика,2007.–128 с.

[14] Жамбю М. *Иерархический кластер-анализ и соответствия*. Изд. 2-е доп. / М. Жамбю — М.: Финансы и статистика, 2002.— 345 с.

[15] *Представление и использование знаний* / Под. Ред. Х. Уэно, М. Исидзука – М.: Мир, 1999. – 220 с.

[16] Muller, J.-A., Lemke, F. *Self-Organizing Data Mining. An Intelligent Approach to Extract Knowledge from Data*. Berlin, Dresden, 1999. – 225 p.

[17] Ивахненко А.Г. *Долгосрочное прогнозирование и управление сложными системами*. – К: Техніка, 1975. – 312 с.

[18] Себер Дж. *Линейный регрессионный анализ*. Изд. 3-е допол. /Дж Себер - М.: Мир, 2005.- 475 с.

[19] Дубровский С.А. *Прикладной многомерный статистический анализ*. Изд. 2-е / С.А. Дубровский – М.: Финансы и статистика, 2002. – 236 с.

[20] Елисеева И. И., Юзбашев М. М. *Общая теория статистики* / Под ред. чл.-корр. РАН И. И. Елисеевой. - М.: Финансы и статистика, 2006. – 368 с.

[21] Кондрашина Е.Ю., Литвинцева Л.В., Поспелов Д.А. *Представление знаний о времении пространстве в интеллектуальных системах* / Под ред. Д.А. Поспелова. - М.: Наука,1999. – 328 с.

[22] Фор А. *Восприятие и распознавание образов*. - М., Машиностроение, 1989. – 302 с.

[23] Раушенбах Г.В. *Меры близости и сходства* // Анализ нечисловой информации в социологических исследованиях. М.: Наука, 1985. – С. 169-203.

УДК 004.681.3

Хорошко В.О., Браіловський М.М. Моніторинг кібератак.
Анотація. *На сьогоднішній день виявлення кібератак є вельми актуальним завданням. Для цієї мети використовується моніторинг мереж. Причому, при цьому виникає необхідність оперативної аналітичної обробки інформації, що вимагає застосування методів інтелектуального аналізу даних. Інтелектуальний аналіз даних допомагає вилучення знань з отриманих даних. Мета притрансформаційних змін інтелектуального аналізу даних до вирішення завдань моніторингу кібернетичних атак - отримання раніше невідомих, нетривіальних, доступних для інтерпретації процесів знань, закономірностей в моніторингу, тобто - даних, корисних для підтримки прийняття рішень. Невід'ємною частиною системи, що розпізнає є навчання, що має кінцевою метою формування еталонних описів класів, форма яких визначається способом їх використання у вирішальних правилах, а також вибір інформаційних ознак для розпізнавання цих еталонних класів. Під час написання даної роботи зроблена спроба викласти в певній логічній послідовності основні аналітичні методи розпізнавання кібернетичних атак в сучасних умовах кібернетичної війни з урахуванням моніторингу інформаційного середовища. Наведено перелік факторів, що підтверджують доцільність застосування методів розпізнавання образів для аналізу даних моніторингу атак. Крім того, розглянуті заходи подібності, які використовуються в алгоритмах ранжирування і кластеризації кібератак. Показано, що доцільність їх застосування залежить від конкретних завдань.*
Ключові слова: *кібератака, моніторинг мереж, методи інтелектуального аналізу, методи розпізнавання образів, системи кіберзахисту.*

Khoroshko V.O., Brailovskyi M.M. Cyber attack monitoring.
Abstract. *To date, the detection of cyberattacks is a very important task. Network monitoring is used for this purpose. Moreover, there is a need for rapid analytical processing of information, which requires the use of methods of data mining. Data mining helps to extract knowledge from acquired data. The purpose of applying data mining to solving problems of monitoring cybernetic attacks is to obtain previously unknown, non-trivial, understandable processes of knowledge, patterns in monitoring, i.e., data useful for supporting decision-making. An integral part of the recognition system is training, which has the ultimate goal of forming reference class descriptions, the form of which is determined by the way they are used in decision rules, as well as the choice of information features for recognizing these reference classes. During the writing of this paper, an attempt was made to set out in a certain logical sequence the main analytical methods for recognizing cyberattacks in modern conditions of cyber warfare, taking into account the monitoring of the information environment. The list of factors confirming expediency of application of methods of recognition of images for the analysis of data of monitoring of attacks is given. In addition, similarity measures used in cyberattack ranking and clustering algorithms are examined. It is shown that the expediency of their application depends on specific tasks.*
Keywords*: cyberattack, network monitoring, methods of intellectual analysis, methods of pattern recognition, cyber defense systems.*

Хорошко Володимир Олексійович, д.т.н., професор, професор кафедри безпеки інформаційних технологій Національного авіаційного університету.
Хорошко Владимир Алексеевич, д.т.н., профессор, профессор кафедры безопасности информационных технологий Национального авиационного университета.
Khoroshko Volodymyr, Doctor of Technical Sciences, Professor, Professor of the Department of Information Technology Security of the National Aviation University.

Браіловський Микола Миколайович, к.т.н., доцент, доцент кафедри кібербезпеки та захисту інформації Київського національного університету імені Тараса Шевченка.
Браиловский Николай Николаевич, к.т.н., доцент, доцент кафедры кибербезопасности и защиты информации Киевского национального университета имени Тараса Шевченко.
Brailovskyi Mykola, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Cyber Security and Information Protection of the Taras Shevchenko National University of Kyiv.

Отримано *22 березня 2021 року*, затверджено редколегією *19 квітня 2021 року*