Разработка подхода к снижению размерности пространства признаков угроз в интеллектуальных системах обеспечивания информационной безопасности

М. А. Пеливан, email: witcher89158779996@uandex.ru
С. А. Будников, email: buser@bk.ru

ФАУ «Государственный научно-исследовательский испытательный институт проблем технической защиты информации ФСТЭК России»

Аннотация. *Разработан подход к снижению размерности пространства признаков угроз в интеллектуальных системах обеспечивания информационной безопасности, основанный на использовании кластерного анализа и критерия принятия решений Сэвиджа, и позволяющий уменьшить трудоемкость решения задачи оценки защищенности объектов защиты, а также эффективности множества мер защиты. Продемонстрирована работа данного алгоритма на примере сведений об угрозах безопасности информации, содержащихся в банке данных угроз ФСТЭК России.*

Ключевые слова: *безопасность информации, интеллектуальные системы, кластеризация, критерий Сэвиджа, снижение размерности, угрозы.*

Введение

В настоящее время происходит активное развитие и внедрение компьютерных технологий во многих аспектах нашей жизни, в следствии чего приходится уделять всё больше внимания организации качественной защиты информации. Одновременно с этим с каждым годом увеличивается возможный ущерб вследствие реализации компьютерных атак, в связи с чем приходится уделять всё больше внимания организации качественной защиты информации (ЗИ).

В системах ЗИ все больше применяются интеллектуальные системы обеспечивания информационной безопасности такие как системы мониторинга событий безопасности (SIEM), системы управления процессами информационной безопасности (SGRC), системы обнаружения целевых атак на конечных точках сети, системы обнаружения и предотвращения вторжений (IPS/IDS) и другие [1]. В основе функционирования перечисленных средств ЗИ лежит работа с большими объемами данных и при этом рост обрабатываемых данных превышает рост вычислительной мощности аппаратных устройств, что

затрудняет как работу интеллектуальные системы обеспечивания информационной безопасности, так и оценку эффективности применяемых мер защиты в целом.

В связи с вышесказанным появляется необходимость в упрощении обработки больших данных, в частности в снижении размерности рассматриваемых признаков данных путем выделения отдельных групп, каждая из которых обладает своей совокупностью наиболее значимых признаков и имеет меньшую размерность.

Целью статьи является разработка методики снижения размерности пространства возможных условий и последствий реализации угроз безопасности информации (далее – снижение размерности).

В данной статье будет проводиться применение разработанного подхода к снижению размерности пространства признаков угроз в интеллектуальных системах обеспечивания информационной безопасности на основе банка данных угроз безопасности информации (БДУ) ФСТЭК России [2]. Информация, содержащаяся в БДУ, позволяет проводить анализ угроз безопасности информации (УБИ) и условий их возникновения на основе: описания УБИ, источников угроз (потенциала нарушителя), объектов воздействия и последствий реализации угрозы.

Однако сама процедура анализа значительно затрудняется из-за большого количества признаков реализации угроз и их сочетаний. На текущий момент БДУ содержит 217 угроз и 27843 уязвимости, 135 объектов воздействия, 6 типов нарушителей (источников угрозы), а также 7 видов последствий реализации угрозы.

Хочется отметить, что при постоянно растущих объемах данных, относящихся к сфере обеспечения информационной безопасности, особое значение приобретает наука о данных (*Data Science*). Именно она занимается созданием и развитием методов автоматизированной обработки больших объемов данных, в том числе и методы, используемые при проведении снижения размерности данных [3]. Одним из таких методов является кластерный анализ, позволяющий производить автоматизированное (автоматическое) формирование групп из множества объектов, получивший широкое распространение в настоящее время [4-5]. Метод предназначен для формирования однородных групп (кластеров) из множества рассматриваемых объектов и характерных для них признаков. Являясь многомерным статистическим методом, кластерный анализ хорошо работает с большим количеством исходных данных. Рассмотрим его более подробно.

В данной работе выбран один из наиболее распространенных алгоритмов кластеризации – алгоритм (метод) *k*-средних [6]. Этот

алгоритм производит формирование k кластеров с наибольшим возможным расстоянием между ними, т.е. покоординатные средние всех кластеров должны максимально отличаться друг от друга.

1. Обоснование подхода к снижению размерности

Обоснуем подход к снижению размерности. Процесс снижения размерности состоит из 4 этапов (рис. 1):

1. Определение исходных данных (объектов кластеризации);
2. Определение количества формируемых кластеров по субтрактивному алгоритму кластеризации;
3. Проведение нескольких итераций процедур кластеризации методом k-средних и выбор наилучшего результата кластеризации;
4. Анализ полученных кластеров.

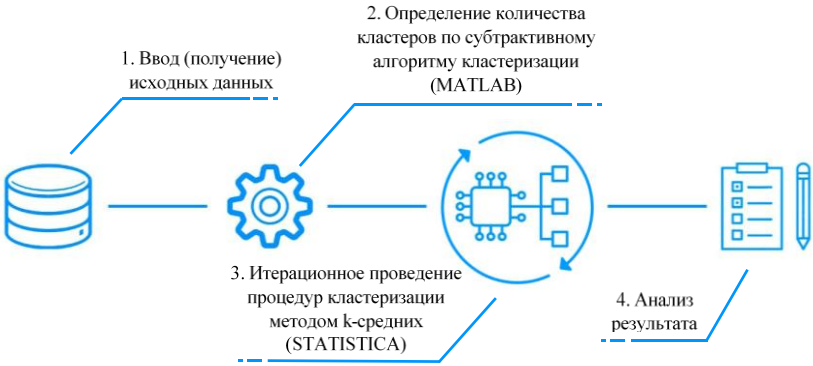

Рис. 1. Общая схема снижения размерности

На первом этапе снижения размерности определяются объекты кластеризации. Для выбранных объектов выбирается критерий (критерии) кластеризации, а также выделить набор признаков, на основании которого и будет осуществляться кластеризация.

На втором этапе определяется количество формируемых кластеров. Одной из особенностей алгоритма k-средних является необходимость самостоятельного определения количества k формируемых кластеров и использование данного значения в качестве исходных данных, что составляет дополнительную проблему. Для определения количества формируемых кластеров выбран метод «горной кластеризации» (субтрактивная кластеризация). Данный вычислительный метод

позволяет определить количество кластеров, не основываясь на интуитивных и теоретических соображениях [7].

Третий этап – кластеризации методом *k*-средних. Как и определение количества формируемых кластеров, сама процедура кластеризации проводится отдельно для каждого критерия кластеризации. Стоит отметить, что выбранный высокоточный алгоритм кластеризации k-средних обладает высокой чувствительностью к выбору начального приближения (центров кластеров), в связи с чем необходимо проведение нескольких процедур кластеризации с использованием различных алгоритмов выбора начальных центров кластеров и последующим определением наиболее подходящего алгоритма для выбранных исходных данных [4].

На четвертом этапе, имея сформированные кластеры, необходимо провести анализ состава всех кластеров и определить общие свойства объектов, находящихся в одном кластере, для этого проводится разделение всех признаков на значимые и незначимые. В данной работе определение значимости признаков будет основано на анализе средних значений и дисперсии отношения количества объектов с данным признаком к общему количеству объектов (далее – отношение количества), а принятие решение о значимости признаков будет проводиться на основе критерия минимального риска Сэвиджа [8]. Данный критерий позволяет выбрать в качестве значимого тот признак реализации УБИ, которому соответствует наибольшее среднее значение отношения количества объектов при наименьшем значении дисперсии отношения количества объектов. Процедура вычисления значений критерия Сэвиджа по формуле (1).

$$S = \min_{i \in N_m} \left\{ \max_{j \in N_n} \left\{ \max_{i \in N_m} a_{ij} - a_{ij} \right\} \right\} \quad (1)$$

где m – число признаков реализации УБИ; n – число оцениваемых характеристик признаков реализации УБИ; a_{ij} – значение j-ой характеристики для i-го признака реализации УБИ; S – критерий Сэвиджа.

2. Реализации разработанной методики снижения размерности

На первом этапе определены исходные данные (объекты кластеризации), используемые для проведения кластеризации. Кластеризация будет проводиться в 2-х пространствах на основе двух критериев (далее – критерий кластеризации) и соответствующих им признаков:

1 критерий кластеризации. Последствия реализации угроз:

– нарушение конфиденциальности, целостности и доступности (КЦД);
– нарушение конфиденциальности и целостности (КЦ);
– нарушение конфиденциальности и доступности (КД);
– нарушение целостности и доступности (ЦД);
– нарушение конфиденциальности (К);
– нарушение целостности (Ц);
– нарушение доступности (Д).

2 критерий кластеризации. Потенциал нарушителя:

– внешний нарушитель с высоким потенциалом (ВншНВП);
– внешний нарушитель со средним потенциалом (ВншНСП);
– внешний нарушитель с низким потенциалом (ВншННП);
– внутренний нарушитель с высоким потенциалом (ВнтНВП);
– внутренний нарушитель со средним потенциалом (ВнтНСП);
– внутренний нарушитель с низким потенциалом (ВнтННП).

Для каждого критерия кластеризации выполняется самостоятельный процесс снижения размерности.

После этого, на втором этапе, произведена субтрактивная кластеризация, реализованную в *Matlab* в виде функции *subclust* [9]. Выполнение субтрактивной кластеризации для обоих критериев кластеризации, «Последствия реализации угроз» и «Потенциал нарушителя», позволило определить количество формируемых кластеров, равное пяти.

Определив значения количества кластеров, осуществляется снижение размерности путем проведения кластеризации методом *k*-средних.

Данная кластеризация производилась с использованием программного обеспечения *STATISTICA*, являющегося набором инструментов для анализа данных, визуализации, прогнозирования, нейросетевых вычислений, *data mining* и контроля качества [10].

Как было отмечено ранее, метод кластеризации *k*-средних обладает высокой чувствительностью к выбору начального приближения (центров кластеров). Поэтому для каждого критерия было осуществлено последовательно три процедуры кластеризации с различными способами выбора начальных центров кластеров:

1. максимизация начальных расстояний между кластерами;
2. сортировка расстояний и выбор наблюдений на постоянных интервалах;
3. выбор первых n (число кластеров) наблюдений.

Первичный анализ результатов кластеризации позволил выделить, для критерия кластеризации «Последствия реализации угроз» процедуру выбора начальных центров кластеров «Максимизация начальных расстояний между кластерами», а для критерия кластеризации «Потенциал нарушителя» – «Сортировка расстояний и выбор наблюдений на постоянных интервалах», как процедуры, позволяющие добиться большей однородности объектов в формируемых кластерах. В результате чего, кластеризация по каждому критерию кластеризации будет осуществляться на основе соответствующих процедур выбора начальных центров кластеров.

Графическое представление результатов кластеризации объектов воздействия по критерию кластеризации «Последствия реализации угроз» представлено на рис. 2.

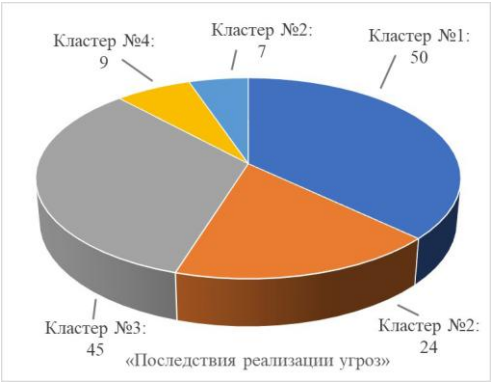

Рис. 2. Результат кластеризации объектов воздействия по критерию кластеризации «Последствия реализации угроз»

Графическое представление результатов кластеризации объектов воздействия по критерию кластеризации «Потенциал нарушителя» представлен на рис. 3.

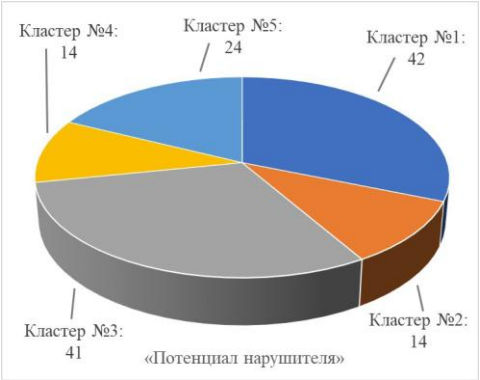

Рис. 3. Результат кластеризации объектов воздействия по критерию кластеризации «Потенциал нарушителя»

На четвертом этапе, определяются значимые признаки кластеров на основе анализа статистических характеристик. Принятие решения о значимости признака проводится на основе анализа средних значений и дисперсий признаков по критерию Сэвиджа.

Результаты вычислений критерия Сэвиджа для кластеров, сформированных на основе признаков кластеризации «Последствия реализации угроз» и «Потенциал нарушителя», представлены в табл. 1 и табл. 2 соответственно.

Таблица 1

Значения критерия Сэвиджа для кластеров, сформированных на основе признака кластеризации «Последствия реализации угроз»

Последствия реализации угроз / Номер кластера	Нарушение КЦД	Нарушение КЦ	Нарушение КД	Нарушение ЦД	Нарушение К	Нарушение Ц	Нарушение Д
Кластер № 1	0,5	0,36	0,5	0,5	0	0,48	0,06
Кластер № 2	0,75	0,79	0,46	0,06	0,69	0,79	0,79
Кластер № 3	0,04	1	0,98	0,96	1	1	1
Кластер № 4	0,44	0,89	0,78	0,67	1	0,28	1
Кластер № 5	0,3	0,86	0,86	0,71	0,6	0,86	0,3

Таблица 2

Значения критерия Сэвиджа для кластеров, сформированных на основе признака кластеризации «Потенциал нарушителя»

Потенциал нарушителя / Номер кластера	ВншНВП	ВншНСП	ВншННП	ВнтНВП	ВнтНСП	ВнтННП
Кластер № 1	0,35	0,16	0,52	0,47	0,52	0
Кластер № 2	1	0,71	0,22	1	1	1
Кластер № 3	0,95	0,8	0,16	1	0,88	0,16
Кластер № 4	0,86	1	1	1	0,13	1
Кластер № 5	0,96	0,08	0,96	1	0,08	0,92

3. Анализ полученных результатов

Таким образом, выбирая минимальные значения критерия Сэвиджа результатов кластеризаций объектов воздействия в пространствах «Последствия реализации угроз» и «Потенциал нарушителя», можно описать каждый сформированный кластер с точки зрения общих (значимых) признаков объектов, включенных в них:

1. По критерию «Последствия реализации угроз»:

– кластер № 1 состоит из объектов, для которых характерны нарушения конфиденциальности и нарушения доступности;
– кластер № 2 состоит из объектов, для которых характерны нарушения целостности и доступности;
– кластер № 3 состоит из объектов, для которых характерны нарушения конфиденциальности, целостности и доступности;
– кластер № 4 состоит из объектов, для которых характерны нарушения конфиденциальности, целостности и доступности, а также нарушения целостности;
– кластер № 5 состоит из объектов, для которых характерны нарушения конфиденциальности, целостности и доступности, а также нарушения доступности.

2. По критерию «Потенциал нарушителя»:

– кластер № 1 состоит из объектов, для которых характерны нарушения, совершаемые внешними нарушителями со средним потенциалом и внутренними нарушителями с низким потенциалом;

- кластер № 2 состоит из объектов, для которых характерны нарушения, совершаемые внешними нарушителями с низким потенциалом;
- кластер № 3 состоит из объектов, для которых характерны нарушения, совершаемые внешними нарушителями с низким потенциалом и внутренними нарушителями с низким потенциалом;
- кластер № 4 состоит из объектов, для которых характерны нарушения, совершаемые внутренними нарушителями со средним потенциалом;
- кластер № 5 состоит из объектов, для которых характерны нарушения, совершаемые внешними нарушителями со средним потенциалом и внутренними нарушителями со средним потенциалом.

Заключение

В данной работе был разработан подход к снижению размерности пространства признаков угроз в интеллектуальных системах обеспечивания информационной безопасности. Разработанный подход включает в себя четыре этапа и основывается на циклическом проведении кластерного анализа с последующей оценкой результатов с применением критерия принятия решений Сэвиджа. В результате применения разработанного подхода для сведений об угрозах безопасности информации, содержащихся в банке данных угроз ФСТЭК России было проведено снижение размерности пространства возможных условий и последствий реализации угроз безопасности информации с 5760 элементов до 25.

Применение разработанного подхода к снижению размерности пространства признаков реализации УБИ актуально при проведении оценки эффективности и обосновании множества мер защиты информации от УБИ. Несмотря на возможную потерю в точности оценки эффективности множества мер защиты информации от УБИ, применение данного подхода позволяет уменьшить трудоемкость и как следствие, материальные затраты, на решение задачи оценки эффективности множества мер защиты информации от УБИ.

Разработанный подход к снижению размерности пространства признаков реализации УБИ позволил выделить значимые (типовые) признаки реализации УБИ по признакам кластеризации «Последствия реализации угроз» и «Потенциал нарушителя», однако для дальнейшего уменьшения трудоемкости решения задачи оценки эффективности и обоснования множества мер защиты информации от УБИ требуется выделение типовых объектов воздействия УБИ. Поэтому дальнейшая

работа будет направлена на обоснование и типизацию объектов воздействия УБИ.

Список литературы

1. Популярные категории средств защиты информации [Электронный ресурс]. – Режим доступа : https://www.anti-malware.ru/security/ свободный. Загл. с экрана – Яз. рус.
2. Банк данных угроз безопасности информации [Электронный ресурс]. – Режим доступа : https://bdu.fstec.ru/ свободный. Загл. с экрана – Яз. рус.
3. Силен, Д. Основы Data Science и Big Data. Python и наука о данных / Д. Силен, А. Мейсман [и др.]. – СПб.: Питер, 2018. – 336 с.
4. Пеливан, М.А. Реализация алгоритмов кластеризации в среде Matlab / М.А. Пеливан // Сборник научных работ III Международной научно-практической конференции «Актуальные вопросы науки, технологии и производства», г. Санкт-Петербург, 21-22 ноября: / – СПб: Издательство Международный Союз ученых, 2014. – С. 134-139.
5. Будников, С.А. Вероятностные характеристики поиска признаков компьютерной атаки в централизованной системе защиты информации / С.А. Будников, И.А. Андреещев // Материалы международной научно-практической конференции «Информационные технологии. Проблемы и решения» №1-2 – Уфа: Издательство Уфимский государственный нефтяной технический университет, 2015. – С. 316-320.
6. Миркин, Б.Г. Методы кластер-анализа для поддержки принятия решений: обзор : препринт WP7/2011/03 / Б. Г. Миркин ; Национальный исследовательский университет «Высшая школа экономики». – М. : Изд. дом Национального исследовательского университета «Высшая школа экономики», 2011. – 88 с.
7. Заде, Л.А. Кластеризация и кластер / Л.А. Заде, С. Рао [и др.]. – М.: Мир, 1980. – 383 с.
8. Богоявленский, С.Б. Теоретические и практические аспекты принятия решений в условиях неопределенности и риска : учеб. пособие / С. Б. Богоявленский. - СПб. : Изд-во С.-Петерб. гос. экон. ун-та, 2014. - 118 с.
9. Дьяконов, В.П. MATLAB. Полный самоучитель / В. П. Дьяковнов ; – М.: ДМК Пресс, 2012. – 768 с.
10. Боровиков, В.П. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов / В. П. Боровиков – СПб.: Питер, 2003. – 688 с.