МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ВИНИТИ РАН)

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 1. ОРГАНИЗАЦИЯ И МЕТОДИКА ИНФОРМАЦИОННОЙ РАБОТЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г. **№ 2** **Москва 2020**

ОБЩИЙ РАЗДЕЛ

УДК 004.021.056:001.103–047.44

О.В. Сюнтюренко

Использование методов аналитической постобработки данных для защиты ресурсов в системах коллективного пользования*

Исследуются теоретические и прикладные аспекты использования методов аналитической постобработки данных с помощью многомерного анализа данных, для защиты ресурсов в системах коллективного пользования. Рассматриваются новые подходы, алгоритмы и процедуры этого процесса на основе регистрационной статистики и многомерного анализа данных, позволяющие противодействовать реализации неявных, косвенных методов несанкционированного доступа (или иных действий) к информации. Предлагается методика оценки качества контролируемых показателей, а также стационарности состояния системы показателей, характеризующих «образ» пользователя в системе. Анализируются вопросы восприятия результатов постобработки данных лицом, принимающим решения (администратором службы безопасности). Представлена методика графической визуализации результатов регистрационно-аналитической обработки зафиксированных данных.

Ключевые слова: *аналитическая постобработка данных, система коллективного пользования, информационные ресурсы, анализ данных, риски и угрозы, несанкционированный действия, информационная безопасность, визуализация*

DOI: 10.36535/0548-0019-2020-02-1

* Статья подготовлена в рамках работ по гранту РФФИ № 20-07-00014 «Разработка методологии использования наукометрических данных для решения задач целеполагания, прогнозирования и управления научными исследованиями».

ВВЕДЕНИЕ

Лавинообразный рост цифровой среды, развитие телекоммуникационной инфраструктуры, широкое применение современных информационных технологий – все это потенциально создает предпосылки возникновения таких угроз, как утечка, хищение, искажение, подделка, копирование и блокирование информации и, как следствие – экономический, экологический, социальный и другие виды ущерба [1–7]. В разных странах регулярно регистрируются многочисленные попытки несанкционированного проникновения в информационные системы органов государственной власти и управления, факты кражи и компрометации экономической и финансовой информации, программного обеспечения систем электронных платежей и т.д. Уровень развития оперативных средств защиты системного программного обеспечения, особенно в современных СУБД типа *Oracle*, позволяет говорить о достаточно надежной защите информационных ресурсов от непосредственных попыток несанкционированного доступа. Однако возможна реализация неявных, косвенных методов несанкционированного доступа (или иных действий) к информации. Как правило, эти методы базируются на определенной априорной осведомленности потенциального нарушителя о характеристиках средств оперативной защиты и возможности логического вывода интересующих данных при применении продуманных стратегий работы с базами данных.

В корпоративных системах коллективного пользования (банки, научно-исследовательские организации, медицинские центры, промышленные корпорации, федеральная налоговая служба, госавтоинспекция и другие правительственные структуры) актуализируются риски несанкционированного использования информационных ресурсов, имеющие свою специфику. Следует отметить, что по оценкам *International Data Corporation* (IDC), цифровая корпоративная информация ежегодно растет на ~40% [6]. Особую напряженность в Социуме вызывают факты хищения и несанкционированного использования персональных данных, которые в последние годы имеют устойчивую тенденцию роста. В подавляющем большинстве случаев эти факты являются следствием противоправных, несанкционированных действий кадровых сотрудников организаций, прежде всего связанных с эксплуатацией, поддержкой и развитием компьютерных систем. Человеческий фактор представляется наиболее слабым и уязвимым звеном при решении задач обеспечения безопасности информационных систем. Далее мы рассмотрим подходы и методы противодействия неявным, скрытым несанкционированным действиям персонала с использованием методов постобработки данных на основе регистрационной статистики и многомерного анализа данных [8].

МЕТОДИКА ОЦЕНКИ КАЧЕСТВА ПОКАЗАТЕЛЕЙ УЧЕТА И КОНТРОЛЯ ПОДСИСТЕМЫ АНАЛИТИЧЕСКОЙ ПОСТОБРАБОТКИ ДАННЫХ

Регистрационно-аналитические средства подсистемы постобработки могут использоваться в двух основных направлениях – в плане защиты информации в локальных интерактивных информационных системах коллективного пользования (СКП) и в плане оптимизации использования информационных и вычислительных ресурсов СКП. Сформулируем принципы разработки и функционирования подсистемы аналитической постобработки регистрационной информации.

Первый принцип – заключается в автоматическом принятии решений о доступе к информации или о автоматической выдаче, в определенных ситуациях, рекомендаций лицу, ответственному за принятие решения. Он состоит в том, что регистрационно-аналитическая подсистема должна отслеживать квазиустойчивые параметры. Рассмотрим пример. Для случайной величины, распределенной по нормальному закону, квазиустойчивыми параметрами являются, например, математическое ожидание и дисперсия. Если вдруг они значимо изменяются, то это означает, что возникли причины, повлиявшие на характер распределения. Таким образом, если регистрационно-аналитическая подсистема определила, что произошло значимое изменение хотя бы одного параметра, то она вырабатывает санкционные процедуры и/или сообщение об этой ситуации для лица, принимающего решения. В качестве параметров такого типа могут использоваться значения определенных показателей или обобщенных факторов, например центроиды кластеров, частоты появления тех или иных признаков, тренды показателей и т.д.

Второй принцип состоит в модульности и программной независимости регистрационно-аналитических средств от программного обеспечения защищаемой системы. Сопряжение этих элементов обеспечивается программами, которые записывают информацию, поставляемую штатными средствами регистрации (на внешнем носителе в определенном формате). Остальные модули регистрационно-аналитической подсистемы работают только с этими форматированными данными. В силу значительного разнообразия состава решаемых функциональных задач, программно-технических средств, структур построения реальных систем коллективного пользования, не представляется возможным разработать унифицированные средства аналитической постобработки. Однако в настоящей статье представлен подход к реализации концептуального проектного решения регистрационно-аналитической подсистемы на основе использования методов многомерного анализа данных [9]. Виртуальная модель гипотетической подсистемы включает в себя комплекс функционально-ориентированных программных модулей, компонуемых со штатными средствами СКП в соответствии с целевыми задачами защиты информации.

Подсистема постобработки обеспечивает учет деятельности пользователя как в ходе сеанса, так и в течение более длительного периода его работы в рамках системы. Для формирования «образа» пользователя базы данных (банка данных, электронной библиотеки, архива) может использоваться ряд показателей, включающих, например, такие параметры, как: 1) число запросов (общее, за сеанс); 2) промежуток времени между запросами; 3) время между поступлениями разных задач (частота сменяемости задач); 4) число ошибок при идентификации; 5) число

нарушений полномочий; 6) время работы процессора для одного задания; 7) число операций чтения; 8) число операций корректировки; 9) число выполненных запросов; 10) число технических сбоев, приходящихся на задачу; 11) число записей (файлов), к которым пользователь обратился за время наблюдения.

Все эти показатели выбираются из системы средств регистрации. Наличие массива данных регистрационных программных средств позволяет реализовать следующие свойства и функциональные задачи подсистемы аналитической постобработки:

- отслеживать псевдоустойчивые параметры пользователя и определять значимость их значений по сравнению с базовыми временными интервалами;
- синтезировать показатели, описывающие информационное поведение пользователей на основе данных, предоставляемых штатными средствами регистрации;
- сжимать и подготавливать информацию к долговременному хранению;
- автоматически приводить в действие механизм принятия санкций и/или формировать рекомендации администратору службы безопасности.

Для выполнения этих функций подсистема аналитической постобработки должна реализовывать ряд статистических методов (процедур), позволяющих решать указанные задачи. Одна из важнейших задач – это формирование представительной системы показателей, характеризующих поведение пользователя. Учитывая значительный объем регистрируемых показателей, решение данной задачи состоит из двух частей:

1. Фильтрация малозначащих показателей. Необходимость фильтрации объясняется тем, что среди показателей могут встречаться равномерно распределенные, которые не влияют на функционирование системы и могут быть исключены из рассмотрения и анализа. Для проверки на равномерность может быть предложен такой метод: проверяется статистическая гипотеза о равенстве средних двух совокупностей, на которые разбита первоначальная совокупность. Они характеризуются средними $\overline{z}$, и дисперсиями σ_z^2, σ_u^2. Выдвигается гипотеза, что эти средние равны, т. е. $H_0 : \overline{z} = \overline{u}$. Для проверки этой гипотезы из каждой совокупности производится выборка: из первой – объемом a_1, в результате получаются $\overline{z^*}$ и G_z^2, из второй – объемом a_2, в результате получаются $\overline{u^*}$ и G_u^2, где G_z^2 и G_u^2 – дисперсии в двух выборках. По этим данным необходимо проверить основную гипотезу, для чего используется статистика:

$$\Phi = \left(\overline{z^*} - \overline{u^*}\right) \cdot \sigma^{-1} \left(\overline{z^*} - \overline{u^*}\right)^{-1}. \qquad (1)$$

Поскольку матожидание $M\left(\overline{z^*}\right) = \overline{z}$ и $M\left(\overline{u^*}\right) = \overline{u}$, то при справедливости гипотезы H_0 будем иметь M(Ф) = 0. Используя свойства дисперсии и полагая выборки независимыми, получим

$$\sigma^2\left(\overline{z^*} - \overline{u^*}\right) = \sigma^2\left(\overline{u^*}\right) + \sigma^2\left(\overline{z^*}\right) = \sigma_z^2 \cdot a_1^{-1} + \sigma_u^2 \cdot a_2^{-1}. \qquad (2)$$

При предположении равенства дисперсий, т.е. $\sigma_z^2 = \sigma_u^2 = \sigma^2$, получаем

$$\sigma^2\left(\overline{z^*} - \overline{u^*}\right) = \sigma^2\left(a_1^{-1} + a_2^{-1}\right). \qquad (3)$$

Подставляя в формулу (1) выражение (3) получим

$$\Phi = \left(\overline{z^*} - \overline{u^*}\right)\left(a_1^{-1} + a_2^{-1}\right)^{-1/2}. \qquad (4)$$

При выборках достаточно большого объема $\overline{z^*}$ и $\overline{u^*}$ распределены нормально, поэтому нормально будет распределена и Ф. Заменив неизвестную дисперсию генеральной совокупности σ^2 ее несмещенной выборочной оценкой

$$G^2 = \left[\sum\left(z_i - \overline{z^*}\right)^2 + \sum\left(u_i - \overline{u^*}\right)^2\right] \cdot (a_1 + a_2 - 2)^{-1} =$$
$$= \left(a_1 G_z^2 + a_2 G_u^2\right) \cdot (a_1 + a_2 - 2)^{-1}$$

приходим к нормально распределенному критерию

$$Y = \left[\left(\overline{z^*} - \overline{u^*}\right) \cdot \left(a_1 G_z^2 + a_2 G_u^2\right)^{-1/2}\right] \cdot \left[(a_1 + a_2 - 2)^{1/2} \cdot \left(a_1^{-1} + a_2^{-1}\right)^{-1/2}\right]. \qquad (5)$$

В общем случае для проверки равномерности могут быть использованы и другие статистические критерии.

Равенство средних является признаком равномерности распределения, и этот показатель исключается из дальнейшего рассмотрения.

2. Выделение групп сильно коррелирующих факторов. После отсеивания равномерно распределенных параметров встает задача выделения групп сильно коррелирующих факторов, каждая из которых может быть заменена одним новым «синтетическим» фактором. Таким образом осуществляется сокращение размерности системы показателей. Для этого мы предлагаем использовать метод главных компонентов, выделяющий некоторые скрытые закономерности, представляющие собой линейные комбинации исходных параметров. Метод главных компонентов заключается в следующем. Некая гипотетическая система описана с помощью n характеристик. Предполагается, что изменения этих характеристик есть реализация случайных величин $x_i, \left(i = \overline{1,n}\right)$, о распределении которых нет никаких предположений, кроме равенства нулю средних. Предполагается также, что система подвергается воздействию таких скрытых, но объективно существующих закономер-

ностей. Отыскание этих закономерностей является *первой задачей* метода главных компонентов. *Вторая задача* – описание изучаемой системы числом главных компонентов *m*, значительно меньшим, чем число первоначально выбранных характеристик *n*. Практические возможности решения этих задач могут быть реализованы в следующих направлениях: 1) сжатие исходной информации; 2) классификация объектов наблюдения; 3) ранжирование объектов или наблюдений по главным компонентам.

В нашем случае метод главных компонентов обладает определенным преимуществом перед другими методами факторного анализа. Он не требует никаких гипотез о переменных, является линейным и аддитивным.

Алгоритмически метод реализуется следующим образом.

Шаг 1 – по заданным наблюдениям рассчитывается корреляционная матрица.

Шаг 2 – матрица корреляций приводится к виду ортогонально подобной ей диагональной, где элементы по главной диагонали есть собственные числа матрицы корреляций.

Шаг 3 – определяется факторная матрица, т.е. матрица, где выявляются нагрузки на выбранные главные компоненты.

Шаг 4 – с помощью итерационной процедуры (алгоритма Хотеллинга) производится вращение матрицы для ориентации ее в направлении главных компонентов, имеющих наибольшие факторные нагрузки.

С точки зрения классификации и отсеивания незначащих показателей методом главных компонентов могут быть рекомендованы три направления:

- если в главные компоненты те или иные показатели входят с высокой факторной нагрузкой (например, более 0,75), то эти показатели связаны между собой и значимы для описания объекта;
- выделяются главные компоненты, имеющие наибольшую факторную нагрузку (например, более 80%), и анализируются определяющие их исходные показатели;
- если те или иные показатели входят в главные компоненты с высокой факторной нагрузкой, но сами главные компоненты не вносят значительного вклада в суммарную дисперсию, то эти параметры не являются важными для описания объекта и могут быть отсеяны.

Для решения задач сокращения размерности исходной системы показателей и формирования их представительной совокупности, имеется ряд методов, в определенной степени связанных с факторным. К таким методам могут быть отнесены:

кластерный анализ – его задача состоит в разбиении множества точек так, чтобы каждая точка принадлежала одному и только одному подмножеству разбиения. При этом в каждом подмножестве разбиения точки лежат плотно друг к другу и являются сходными, в то время как точки, принадлежащие разным подмножествам, разнородны. В качестве меры близости точек в кластерном анализе обычно используются следующие критерии:

а) евклидово расстояние

$$d_{ij}^2 = \sum_{l=1}^{P} \left(x_{il} - x_{jl}\right)^2 \ ;$$

б) взвешенное евклидово расстояние

$$d_{ij}^2 = \sum_{l=1}^{P} \omega_i \left(x_{il} - x_{jl}\right)^2 ;$$

в) расстояние Махалонобиса (оно отличается от евклидова расстояния) тем, что учитывает корреляции между переменными и инвариантно к масштабу)

$$d_{ij}^2 = \left(x_i - x_j\right)^T S^{-1} \left(x_i - x_j\right) ,$$

где S^{-1} – матрица, обратная ковариационной;

г) коэффициент корреляции

$$d_{ij}^2 = \left\{ \left[\left(x_{il} - \overline{x_l}\right)\left(x_{jl} - \overline{x_l}\right) \right] \cdot \left(\sigma_i \cdot \sigma_j\right)^{-1} \right\}^{1/2} .$$

В общем случае задача кластерного анализа сводится к разработке определенного правила или алгоритма, с помощью которого можно осуществлять разбиение точек на группы. При этом остается неизвестно, действительно ли найденные группировки являются наилучшими. В кластерном анализе при построении процедуры группировки используется *R* – коэффициент или коэффициент принадлежности

$$R = 100 \left(\frac{\begin{array}{c}\textit{средний коэффициент корреляции между} \\ \textit{переменными одной группы}\end{array}}{\begin{array}{c}\textit{средний коэффициент корреляции переменных} \\ \textit{этой группы с остальными переменными}\end{array}} \right) \quad (6)$$

При этом не разработано никакого удовлетворительного статистического критерия, который позволял бы оценить приведенное разбиение и принадлежность точки к определенной группе.

анализ образов – это изучение структуры количественных данных, исходя из коэффициентов множественной регрессии. Каждую переменную предлагается оценивать с помощью метода множественной регрессии по остальным (*m*-l) переменным. Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель [10].

В некоторых случаях, в связи с относительно большой вычислительной сложностью изложенных процедур, средства программной поддержки целесообразно использовать в виде стандартных программ в соответствующих пакетах [11, 12].

МЕТОДИКА ОЦЕНКИ СТАЦИОНАРНОСТИ СОСТОЯНИЯ СИСТЕМЫ ПОКАЗАТЕЛЕЙ, ХАРАКТЕРИЗУЮЩИХ ПОЛЬЗОВАТЕЛЯ

Подсистема постобработки данных должна отслеживать квазиустойчивые параметры. Например, для случайной величины, имеющей нормальный закон распределения, такими параметрами являются математическое ожидание и дисперсия. Значительное изменение этих параметров означает, что возникли причины, изменившие характер распределения. В свою очередь, стабильность параметров соответствует определенной модели поведения пользователя. Изменение этой модели может быть вызвано, в частности, попыткой несанкционированного доступа. Таким образом, если регистрационно-аналитическая подсистема определила, что произошло значимое изменение хотя бы одного из таких параметров, то она реализует санкционные процедуры и/или формирует сообщение об этой ситуации для лица, принимающего решения. Отсюда вытекает необходимость оценки стационарности системы показателей, которая позволит определенным образом делать выводы о поведении пользователей. Под стационарностью понимается устойчивость статистических характеристик системы во времени. Поскольку показатели, выбираемые из системных средств защиты, представляют собой реализацию случайных величин, то для проверки на стационарность предлагается использовать гипотезу о сравнении долей признака в двух совокупностях.

Пусть $\frac{m_1}{n_1}$ и $\frac{m_2}{n_2}$ – частности одного и того же признака в двух совокупностях из n_1 и n_2. Гипотезой H_0 является предположение, что обе совокупности представляют собой две выборки из одной генеральной совокупности с некоторой долей признака ρ, а отмеченное расхождение выборочных частностей есть результат случайностей, сопровождающих отбор. При построении критерия проверки различаются большие и малые выборки.

1. Большие выборки. Если n_1 и n_2 – большие числа (примерно более 30), то распределение выборочных частностей будет близко к нормальному с параметрами

$$M\left(m_1 n_1^{-1}\right) = M\left(m_2 n_2^{-1}\right) = \rho,$$

и дисперсиями

$$\sigma^2\left(m_1 n_1^{-1}\right) = \rho(1-\rho)\cdot n_1^{-1},\ \sigma^2\left(m_2 n_2^{-1}\right) = \rho(1-\rho)\cdot n_2^{-1}.$$

Введем для проверки гипотезы статистику

$$\theta = \left(m_1 n_1^{-1}\right) - \left(m_2 n_2^{-1}\right).$$

При справедливости гипотезы H_0 значение θ может лишь случайно отличаться от нуля. Статистика θ также подчиняется нормальному закону с параметрами:

$$M(\theta) = M\left[\left(m_1 n_1^{-1}\right) - \left(m_2 n_2^{-1}\right)\right] =$$
$$= M\left(m_1 n_1^{-1}\right) - M\left(m_2 n_2^{-1}\right) = \rho - \rho = 0;$$

$$\sigma^2(\theta) = \sigma^2\left[\left(m_1 n_1^{-1}\right) - \left(m_2 n_2^{-1}\right)\right] =$$
$$= \sigma^2\left(m_1 n_1^{-1}\right) - \sigma^2\left(m_2 n_2^{-1}\right) = \rho(1-\rho)\left(n_1^{-1} - n_2^{-1}\right).$$

Далее переходим к проверке нулевой гипотезы с помощью двустороннего критерия, который диктуется смыслом проверки. Задавшись уровнем значимости $\propto$, найдем $z_{\propto/2}$ из уравнения

$$\rho\left\{|\theta| < z_{\propto/2}\sigma(\theta)\right\} = 2\Phi(t) = 1 - \propto. \qquad (7)$$

Исходя из уравнения (7) получим критические точки

$$\theta_1 = -z_{\propto/2}\left[\rho(1-\rho)\right]^{1/2}\left(n_1^{-1} + n_2^{-1}\right)^{1/2},$$
$$\theta_2 = z_{\propto/2}\left[\rho(1-\rho)\right]^{1/2}\left(n_1^{-1} + n_2^{-1}\right)^{1/2},$$

где величину ρ заменяем ее оценкой на основании данных двух выборок

$$\rho = (m_1 + m_2)\cdot(n_1 + n_2)^{-1}.$$

Правило проверки: если выборочное значение θ^* лежит в интервале (θ_1, θ_2), то гипотеза H_0 не отклоняется, т.е. расхождение между $\frac{m_1}{n_1}$ и $\frac{m_2}{n_2}$ несущественно; если θ^* окажется вне этого интервала, то H_0 отклоняется, т.е. расхождение между $\frac{m_1}{n_1}$ и $\frac{m_2}{n_2}$ можно рассматривать как значимое.

2. Малые выборки. На практике это более сложный случай. Если n_1 и n_2 – малые числа, то использование нормального распределения для статистики $\theta = \left(m_1 n_1^{-1}\right) - \left(m_2 n_2^{-1}\right)$ становится некорректным. В этом случае пользуются критерием χ^2, с помощью которого оценивается расхождение между теоретическими и выборочными частотами. Критерий χ^2 вычисляется следующим образом: через $\overline{m_1}$

и $\overline{m_2}$ обозначается число элементов, не обладающих признаком *A*. Если это выборки из одной и той же генеральной совокупности с долей признака ρ, то можно определить теоретические частоты ρn_1, $(1-\rho)n_1$, ρn_2, $(1-\rho)n_2$. При этом для ρ принимается оценка $\rho = (m_1 + m_2)\cdot(n_1 + n_2)^{-1}$. Вычисляем χ^2 по формуле

$$\chi^2 = (m_1 - \rho n_1)^2 (\rho n_1)^{-1} +$$
$$+\left[m_1 - (1-\rho)n_1\right]^2 \left[(1-\rho)n_1\right]^{-1} +$$
$$+\left[(m_2 - n_2\rho)\right]^2 (\rho n_2)^{-1} +$$
$$+\left[m_2 - (1-\rho)n_2\right]^2 \left[(1-\rho)n_2\right]^{-1}$$

При этом, так как между четырьмя теоретическими частотами существуют три независимых соотношения, то независимой является только одна величина, т.е. в распределении χ^2 следует учесть одну степень свободы $(\nu = 1)$.

Если нулевая гипотеза (согласно которой обе совокупности есть выборки из одной генеральной совокупности) верна, то расхождение между теоретическими и опытными частностями можно отнести только к случайностям отбора. Поэтому, определив для данного уровня значимости $\propto$ значение χ_0^2, примем решение об отклонении гипотезы H_0, если $\chi^2 > \chi_0^2$, и о не значимости расхождений, если $\chi^2 \le \chi_0^2$.

Такая проверка позволяет определить стационарность представленной системы показателей и делать выводы о поведении пользователей, не проводя объемных расчетов всех показателей.

МЕТОДИКА ВИЗУАЛИЗАЦИИ РЕЗУЛЬТАТОВ РЕГИСТРАЦИОННО-АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ

В системах с постобработкой одной из важнейших задач является восприятие информации лицом, принимающим решения (администратором службы безопасности). Это связано с тем, что данные, выдаваемые системными средствами, имеют большой объем и представлены не в наглядной форме. Одной из задач постобработки является визуализация имеющихся данных, т.е. представление их в форме, удобной для восприятия человеком. Для этой цели используются методы математической статистики (метод главных компонентов, кластер-анализ), которые решают задачу визуализации путем сокращения размерности системы исходных показателей и выявления внутри ее скрытых закономерностей. Большое значение имеют также методы графического представления системы исходных показателей как с помощью обычных графиков, так и с помощью разного рода гистограмм [13].

В различных пакетах прикладных программ используются разные средства визуализации исходной системы показателей. С помощью графиков возможно визуализировать информацию о поведении пользователя в процессе работы с базами данных. Например, может быть дана оценка разнообразия информационных потребностей пользователя.

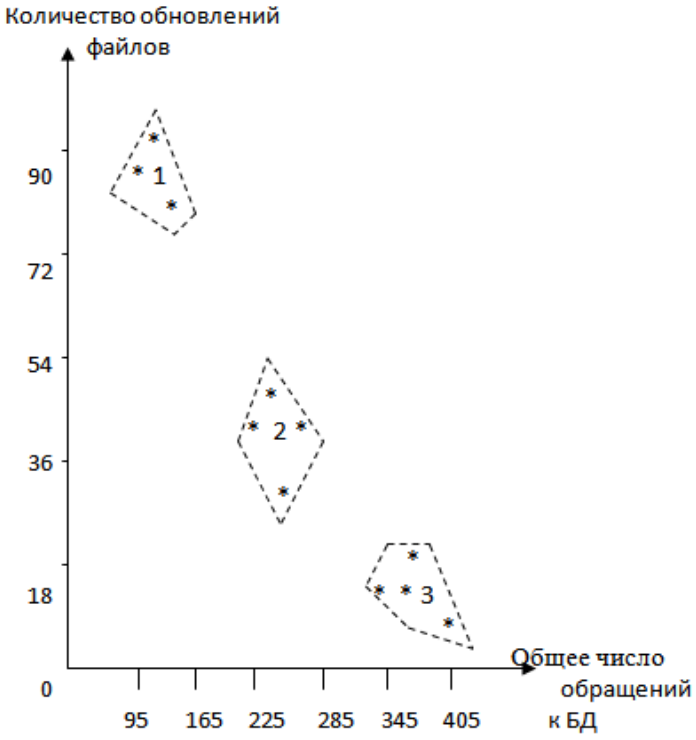

Рис. 1. Визуализация взаимосвязи количества обновлений файлов и общего числа обращений к БД

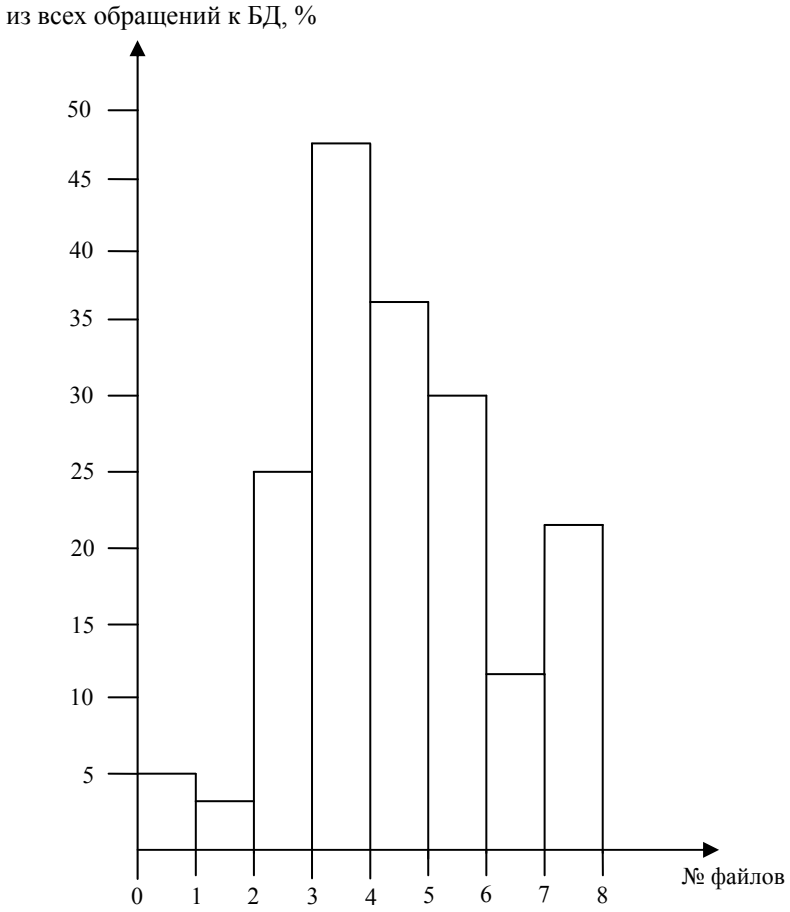

Рис. 2. Распределение по файлам общего количества обращений к БД

Значительный интерес представляет распределение во времени количества обращений к информационным ресурсам системы. Визуализация распределения результатов регистрационно-аналитической обработки информации дает возможность наглядно представлять и группировать пользователей по количеству обращений к базам данных. Такая информация позволяет анализировать взаимность продолжительности работы и ее интенсивность, что может косвенно указывать на попытки доступа к запрещенной информации. Важно также выявить взаимосвязи между количеством обновлений файлов и общим числом обращений к базе данных (рис. 1). Установление такой взаимосвязи позволяет создать «образ» пользователя и по нему делать выводы о способе действий с БД. Например, пользователь, который мало обращается к базе и часто обновляет файлы, может представлять собой потенциального нарушителя. Определенный интерес при анализе деятельности пользователей имеет оценка частоты обращения к файлам (записям) базы данных. Возможность такой оценки связана с тем, что важность файлов, с точки зрения сохранения конфиденциальности (секретности) информации, может быть неодинакова, что можно показать на гистограмме распределения общего количества обращений к базе данных по файлам (рис. 2).

Применение перечисленных методов визуализации результатов аналитической постобработки информации позволяет облегчить анализ накопленной регистрационной информации, повысить эффективность процессов принятия решений администратором службы безопасности организации.

ЗАКЛЮЧЕНИЕ

Несмотря на интенсивное развитие компьютерных средств и информационных технологий, уязвимость современных информационных систем и компьютерных сетей не уменьшается, а имеет тенденцию к нарастанию различного рода рисков и угроз. Основными факторами, способствующими повышению уязвимости информации, являются: рост объемов информации в интегрированных базах данных, представляющей интерес для значительного числа пользователей (как легитимных, так и не имеющих права доступа к ней); развитие телекоммуникационных режимов обработки информации (в сетях Интернет и Интранет); расширение круга пользователей банков данных систем коллективного пользования.

В последнее время в России повышение интереса к проблематике информационной безопасности объясняется главным образом интенсификацией процессов информатизации государственных органов (в том числе МВД, вооруженных сил) [14], банковского и страхового бизнеса, развитием крупных коммерческих структур, а также повышением уровня криминогенной обстановки и террористических угроз в сетевой среде.

Рассмотренные в настоящей статье методы постобработки регистрационной информации позволяют реализовывать выявление скрытых, неявных несанкционированных действий в системах коллективного пользования, осуществляемых сотрудниками организаций (связанных с эксплуатацией, поддержкой и развитием компьютерных систем) и которые не обнаруживаются обычными оперативными программными средствами разграничения доступа к информации. Процедуры и алгоритмы, реализуемые на основе рассмотренных методов многомерного анализа данных, имеют хорошую перспективу применения в различных приложениях в системах компьютерной логики, искусственного интеллекта, в рамках развития аналитических технологий Больших Данных.

В области оперативных средств перспективное направление разработок состоит в создании «безопасных» аппаратных средств, «безопасных» операционных систем (например, на базе *Unix*) и СУБД, т.е. в автоматизации средств защиты на всех уровнях доступа к информационным и вычислительным ресурсам систем коллективного доступа.

СПИСОК ЛИТЕРАТУРЫ

1. Хоффман Л.Дж. Современные методы защиты информации. – М.: Советское радио, 1980. – 263 с.
2. Смолян Г.Л. Сетевые информационные технологии и проблемы безопасности личности // Вестник РФФИ. – 1999. – № 3(17). – С. 63–68.
3. Siountiourenko O. The Problems of Providing Information Security: The Case of Information Infrastructure / ed. Gerhard Banse // Studies in Eastern Europe. Technological and nvironmental Policy. – Berlin, 2007. – P. 161–179.
4. Акопян Д.А., Еляков А.Д. Киберпреступления в информационной инфраструктуре общества (Обзор) // Научно-техническая информация. Сер. 1. – 2009. – № 12. – С. 1-14.
5. Сюнтюренко О.В. Социальные и экономические риски развития информационных технологий // Научно-техническая информация. Сер.1. – 2012. – № 6. – С. 1-5; Syuntyurenko O.V. The Social and Economic Risks of the Development of Information Technologies // Scientific and Technical Information Processing. – 2012. – Vol. 39, № 2. – P. 113-116.
6. Сюнтюренко О.В. Цифровая среда: тренды и риски развития // Научно-техническая информация. Сер. 1. – 2015. – № 2. – С. 1 -7; Syuntyurenko O.V. The Digital Enviroment: The Trends and Risks of Development // Scientific and Technical Information Processing. – 2015. – Vol. 42, № 1. – P. 24-29.
7. Арутюнов В.В. О некоторых результатах приоритетных исследований в области информационной безопасности // Научно-техническая информация. Сер. 1. – 2016. – № 2. – С. 8-13.
8. Сюнтюренко О.В. Теоретические и прикладные аспекты автоматизации процедур многомерного анализа данных // Научно-техническая информация. Сер. 2. – 2018. – № 11. – С. 1-8; Syuntyurenko O.V. Theoretical and Applied Aspects of Automating Multivariate Analysis Procedures // Automatic Documentation and Mathematical Linguistics. – 2018. – Vol. 52, № 6. – P. 275-281
9. Сюнтюренко О.В. Цифровая среда: аналитическая постобработка информации с использованием методов наукометрии и анализа данных // Научно-техническая информация. Сер. 1. – 2019. – № 4. – С. 8-16; Syuntyurenko O.V. Digital Enviroment: Information Analytical Postprocessing Using the Scientometric and Data Analysis Methods // Scientific and Technical Information Processing. – 2019. – Vol. 46, № 2 – P. 59-66.
10. Множественная линейная регрессия. Улучшение модели регрессии. – URL: https://function-x.ru/statistics_regression2.html (дата обращения 12.06.2019).
11. Шарстнев В.Л., Вардомацкая Е.Ю. и др. Пакеты прикладных программ для статистического анализа данных. – URI: https://rep.vstu.by/handle/123456789/1746 (дата обращения 06.06.2019).
12. Величко В.В. Сравнительный анализ статистических пакетов программ. – URL: https://cyberleninka.ru/article/n/sravnitelnyy-analiz-statistiches-kin-paketov-programm (дата последнего обращения 06.03.2019).
13. Клышинский Э.С., Рыасков С.В., Шихов А.И. Обзор методов визуализации многомерных данных. – URL: https://cyberleninka.ru/article/n/obzor-metodov-visualizatsii-mnogomernyh-dannyh (дата последнего обращения 07.03.2019).
14. Национальная Программа «Цифровая экономика Российской Федерации»; утверждена распоряжением Правительства Российской Федерации от 28 июля 2017 г. N 1632-р. – URI: static.government.ru/media/files/9gFM4FHj4PsB79I5v7yLVuPgu4bvR7M0.pdf(дата обращения 27.12.2019).

Материал поступил в редакцию 17.01.20.

Сведения об авторе

СЮНТЮРЕНКО Олег Васильевич – доктор технических наук, профессор, ведущий научный сотрудник ВИНИТИ РАН, Москва
e-mail: olegasu@mail.ru