S.D. Kulik

Efficient Factual Search And Requirements Of The Information Security

Keywords: information security, efficiency, factographicinformation retrieval.
This article is about special methods of information security. These methods are a part of the automated tools of ensuring the information security. It is offered to use a special factographic information retrieval and Automated Factographic Information Retrieval System (AFIRS).

С.Д.Кулик

ЭФФЕКТИВНЫЙ ФАКТОГРАФИЧЕСКИЙ ПОИСК С УЧЕТОМ ТРЕБОВАНИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Данная статья – частичное продолжение работ [1-3], целью которой является дальнейшее исследование (опираясь на [4-6]) эффективности выбранного класса фактографических систем в составе средств обеспечения информационной безопасности.

Есть целый класс информационных систем (ИС), которые принято называть фактографическими системами (ФС). Среди ФС можно выделить более узкий класс систем [1, 2] – «автоматизированные фактографические информационно-поисковые системы» (АФИПС). Нейронные сети [3, 7, 8] позволяют решать достаточно сложные практические задачи, например [8] плохо обусловленные задачи. Анализ эффективности систем показывает, что обеспечение информационной безопасности [1] является очень важной и актуальной задачей. При построении реальных систем с целью повышения эффективности работы автоматизированных средств обеспечения информационной безопасности (АСОИБ) предлагается использовать в составе АСОИБ именно ФС. Можно полагать, что одна из главных задач, решаемых этой ФС, – эффективный фактографический поиск (ФП) [2] с учетом требований информационной безопасности.

Введем следующие частные показатели, связанные с эффективностью защиты информации в данных фактографических информационных системах:

Δ_1 – оценка вероятности неискажения информации от внутреннего злоумышленника;

Δ_2 – оценка вероятности неискажения информации от внешнего злоумышленника;

Δ_3 – оценка вероятности неискажения информации от вредоносного *программного обеспечения* (например, компьютерного вредоносного вируса);

Δ_4 – оценка вероятности неискажения информации от вредоносного *аппаратного обеспечения* (например, закладочных устройств, аппаратных закладок, т.е. устройств в электронной схеме, скрытно внедряемых к остальным элементам);

Δ_5 – оценка вероятности неискажения информации от внешнего вредоносного воздействия на: *аппаратуру*, *программное обеспечение*, *фактографические данные* (например, электромагнитное воздействие).

Будем полагать, что случайные события, соответствующие вероятностям Δ_k, являются независимыми в совокупности событиями. Тогда, опираясь на работы [4, 5, 6], можно полагать, что:

$$\hat{p}_1 = P_1 \prod_{k=1}^{5} \Delta_k = P_1 \Delta_1 \Delta_2 \Delta_3 \Delta_4 \Delta_5 ,$$

$$\hat{p}_2 = P_2 \prod_{k=1}^{5} \Delta_k = P_2 \Delta_1 \Delta_2 \Delta_3 \Delta_4 \Delta_5,$$

где [2]:

P_1– вероятность правильного сравнения двух тождественных объектов по их описаниям (где $(1–P_1)$ определяет вероятность *пропуска цели*);

P_2– вероятность правильного сравнения двух нетождественных объектов по их описаниям (где $(1–P_2)$определяет вероятность *ложной тревоги*).

На практике поисковым блоком ФС сравнение поискового образа запроса (ПОЗ) и поискового образа объекта (ПОО) выполняется с применением алгоритма распознавания образов (например, нейросетевой алгоритм), который характеризуется этими вероятностями P_1, P_2. В целом же само сравнение ПОЗ и ПОО поисковым блоком ФС характеризуется вероятностями $\hat{p}_1$, $\hat{p}_2$, что и позволяет учитывать эффективность защиты информации в фактографических данных.

Идея оценивать информационную безопасность для фактографических систем с помощью показателя p ранее не использовалась. Будем далее полагать, что как-то удалось оценить эту информационную безопасность по отношению к записям фактографической БД (ФБД). На практике злоумышленник может частично разрушить (исказить) содержимое записи данных ФБД, при этом целью ФС является поиск фактографических данных, необходимых для эффективного функционирования АСОИБ.

В случае, когда поисковым блоком с помощью блока распознавания в результате сравнения двух объектов определена их идентичность (тождество), то регистрационный номер ПОО заносится в рекомендательный список (РС). В другом случае, когда поисковым блоком с помощью блока распознавания в результате сравнения двух объектов не определена их идентичность, то регистрационный номер ПОО не заносится в РС. Всего в РС может быть занесено L таких номеров. В процессе сравнения эти поисковые образы объектов поступают из своего поискового массива последовательно один за другим для сравнения поисковым блоком. Поиск прекращается после заполнения РС или после просмотра всей области поиска из N описаний объектов поискового массива, что соответствует стратегии полного поиска [2]. Результатом поиска поискового блока является РС, который либо пуст, либо содержит от 1 до L штук регистрационных номеров (РН) рекомендованных ПОО. Далее под длиной РС будем понимать число содержащихся в нем РН.

В АСОИБ в процессе индексирования возможны ошибки из-за искажения фактографической информации в данных. Поэтому как ПОО, так и ПОЗ, который хранится в поисковом массиве, может содержать ошибки индексирования. В ФС эти ошибки могут приводить в процессе фактографического поиска к *пропускам целей* и *ложным тревогам* и в итоге к ошибочной работы АСОИБ.

Для рассматриваемого класса АСОИБ объекты, чьи описания поступают на вход поискового блока, могут быть описаны, например, тремя группами признаков. Первую и вторую группу составляют признаки, которые поддаются формализованному описанию и могут быть использованы при автоматизированном индексировании объектов и при сравнении объектов (например, нейронной сетью в блоке распознавания) по их описаниям. Первую группу составляют признаки, устойчивые к незначительным искажениям объектов. Вторую группу составляют признаки, чувствительные к незначительным искажениям объектов. Ошибка индексирования этих признаков человеком-оператором или в некоторых случаях автоматом (например, с помощью нейронной сети) сильно зависит от уровня искажения самих объектов. Третью группу составляют признаки, которые не поддаются формализованному описанию и используются только

человеком-оператором (лицом, принимающим решение) лишь при непосредственном сравнении объектов между собой при обработке РС. На основе первой группы признаков строится схема классификации объектов по их описаниям. С ее помощью все описания объектов (ПОО или ПОЗ) разбивают на *K* классов и организуют стратегии поиска. С помощью второй группы признаков строится алгоритм распознавания (сравнения) объектов по их описаниям ПОЗ и ПОО. С его помощью поисковый блок формирует РС для человека-оператора, Далее будем предполагать, что отсутствуют признаки, устойчивые к незначительным искажениям объектов, и поэтому нет схемы классификации объектов.

Объектами ФС в АСОИБ могут быть, например, биометрические портреты заданной группы людей. В АСОИБ обработка запросов на поиск происходит по схеме, показанной на рис.1.

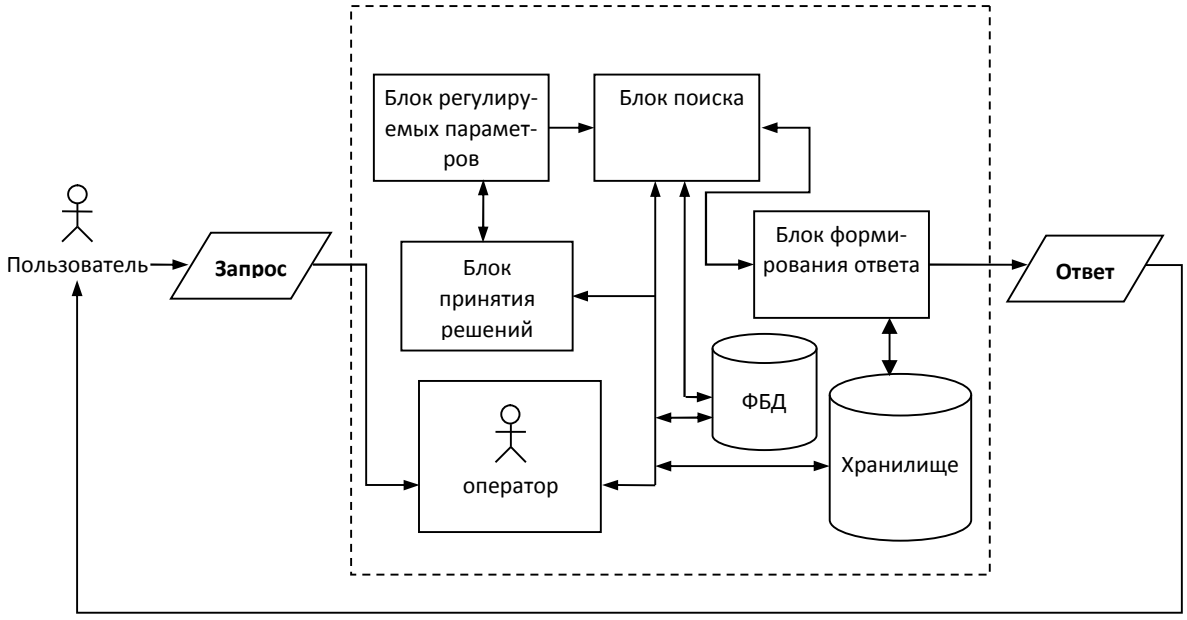

Рис. 1. Краткая схема АСОИБ

На практике специальные средства, которые здесь не рассматриваются, формируют поток запросов в виде объектов-запросов. В АСОИБ эти объекты-запросы поступают оператору для индексирования объекта, где составляется описание объекта в виде ПОЗ. Обычно в простейшем случае ПОЗ содержит перечень значений признаков, которые в дальнейшем используются алгоритмом распознавания в блоке принятия решения. Блок поиска для данного ПОЗ находит в поисковом массиве (ФБД) описания объектов (т.е. ПОО), наиболее похожих на объект-запрос (т.е. ПОЗ). Такая степень «*похожести*» определяется алгоритмом сравнения (распознавания) ПОЗ и ПОО. Далее регистрационные номера (РН) похожих (т.е. рекомендуемых для дальнейшего анализа) объектов заносятся в РС. Так как обычно размер буфера под РС ограничен, то поиск прекращается либо после заполнения РС не более *L* регистрационными номерами объектов, либо после просмотра всей области поиска из *N* штук ПОО в ФБД. Затем сформированный РС, объект-запрос и объекты-хранения из хранилища, чьи регистрационные номера указаны в РС, передаются человеку-оператору на обработку РС. Результирующее (окончательное) решение в ответ на поступивший запрос принимает именно человек-оператор путем анализа (экспертизы) и сравнения объекта-запроса с объектами-хранения, указанными в РС. По результатам обработки РС выдается ответ либо типа

«ДА» (да, есть такой же объект в системе (идентичный объекту-запросу) и по нему имеются такие-то сведения – фактографические данные), либо ответ типа «НЕТ» (нет подобного объекта в системе, идентичного объекту-запросу). Окончательный ответ о результатах поиска передается на выход АСОИБ.

Были проведены исследования и получены аналитические выражения для оценки эффективности фактографического поиска с помощью показателя V_x,:

$$V_x = Pz \cdot S1 + (1 - Pz) \cdot S2 ,$$

где:

$S1$ – вероятность правильного ответа поискового блока на запрос при поиске в области, содержащей тождественный запросу объект (ТЗО);

$S2$ – вероятность правильного ответа поискового блока на запрос при поиске в области, не содержащей ТЗО.

Для АСОИБ с целью получения оценки V_x – вероятности правильного ответа на запрос, как и в работах [2, 4], было введено пространство Ω, где его элементарными событиями являются ответы поискового блока на запрос. В данном случае это пространство событий Ω разделяется на два непересекающихся пространства D и E. Так, пространству D соответствуют события правильного ответа поискового блока на запрос, а пространству E соответствуют события неправильного ответа поискового блока на запрос. Такое разделение пространства Ω позволяет вычислить вероятность появления хотя бы одного события из D. Для АСОИБ именно эту вероятность мы примем за вероятность правильного ответа поискового блока на запрос. Далее аналогично получим, что V_0 – вероятность появления хотя бы одного события из E есть вероятность неправильного ответа поискового блока на запрос.

Из теории вероятностей следует, что события пространства Ω должны составлять полную группу событий, то есть вероятность появления хотя бы одного события из Ω должна быть равна единице, т.е. $V_x + V_0 \equiv 1$.

Введенное пространство D для вероятности $S1$ в зависимости от практических приложений может определяться различными способами. Рассмотрим только один из них. Это пространство D состоит из событий, при которых РН тождественного объекта выдается в РС и при этом может быть выдано от 0 до (L–1) регистрационных номеров объектов, не тождественных запросу.

Также в зависимости от различных практических приложений пространство D для вероятности $S2$ может определяться различными способами. Рассмотрим только один из них: пространство D состоит из событий, при которых в РС выдается от 0 до L регистрационных номеров объектов, не тождественных запросу. По аналогии можно ввести элементарные события пространства Ω_1 для показателя L_x.

Оценка эффективности фактографического поиска

По результатам проведенных исследований были получены для оценки эффективности фактографического поиска в АСОИБ следующие аналитические выражения показателей вероятностей $S1$ и $S2$ [2]:

$$S1 = V1(N, \hat{p}_1, \hat{p}_2, L, \beta_n) = \left\{ \sum_{n=1}^{L} \beta_n + \sum_{n=L+1}^{N} \beta_n \left[\sum_{m=0}^{L-1} C_{n-1}^{m} (1 - \hat{p}_2)^m \, \hat{p}_2^{n-m-1} \right] \right\} \hat{p}_1 ;$$

$$S2 = G4(N, \hat{p}_2, L) = \sum_{m=0}^{N} \left[C_N^m \left(1 - \hat{p}_2\right)^m \hat{p}_2^{N-m} \right].$$

Эти полученные оценки позволяют учесть искажения информации, человека-оператора и оценить влияние алгоритма распознавания на эффективность поискового блока в АСОИБ.

Выполнены необходимые исследования [2] и было установлено для функций $S1$ и $F2$, как они изменяются при изменении значений параметров L, N, P_1, и P_2 . Небольшая часть результатов этих исследований для $S1$ при $\beta_n = \frac{1}{N}$ (где n=1,2,3,...,N), различных N и L представлена в табл. 1, 2.

Таблица 1. Эффективность при N=800, P_1=0.9, P_2=0.85

Показатели эффективности	Значения показателей эффективности								
L	40	100	110	115	118	120	130	140	800
$S1$	0.3	0.75	0.82	0.85	0.86	0.87	0.89	0.9	0.9

Таблица 2. Эффективность при N=500, P_1=0.9, P_2=0.85

Показатели эффективности	Значения показателей эффективности							
L	40	60	70	75	80	90	400	500
$S1$	0.48	0.72	0.82	0.86	0.88	0.89	0.9	0.9

Из табл. 1 следует, что при P_1=0.9, P_2=0.85, $\beta_n = \frac{1}{N}$, при числе ПОО (записей) N=800 и при ограничении на максимально возможную длину РС в 140 регистрационных номеров (L=140) вероятность правильного ответа поискового блока на запрос при фактографическом поиске в области, содержащей ТЗО, составляет $S1$=0.9 для АСОИБ.

Выводы

Кратко для АСОИБ рассмотрен специальный класс информационных систем – фактографических информационных систем. Представлен подход к обеспечению эффективного ФП в АСОИБ. Предложена концепция построения АСОИБ с учетом фактографической системы. Приведено описание варианта возможной схемы для реализации АСОИБ с учетом возможного искажения информации. Намечены важные пути реализации фактографического поиска в АСОИБ. На данном этапе исследования получены результаты, способствующие эффективному решению задачи обеспечения информационной безопасности. Получены аналитические выражения для оценки выбранных показателей эффективности фактографического поиска в АСОИБ. Выполнено исследование этих оценок и выявлены их свойства.

СПИСОК ЛИТЕРАТУРЫ:

1. *Кулик С.Д.* Информационная безопасность АФИПС // Известия Южного федерального университета. Технические науки. 2003. Т. 33. № 4. С. 238.

2. *Кулик С.Д.* Алгоритмы распознавания образов и моделирование автоматизированных фактографических информационно-поисковых систем // Нейрокомпьютеры: разработка и применение. 2007. №2-3. С.67-82.

3. *Кулик С.Д* .Применение нейросетевого подхода в информационных и экспертных системах // Нейрокомпьютеры: разработка и применение. 2007. №2-3. С.93-118.
4. *Вентцель Е.С.* Теория вероятностей. М.: Высшая школа, 2001. 575 с.
5. *Колмогоров А.Н.* Основные понятия теории вероятностей. М.: ФАЗИС, 1998. 144 с.
6. *Feller W.* An introduction to probability theory and its applications, Vol.1, 3nded., John Wiley & Sons. New York, 1968. xvii+509 pp.
7. *Галушкин А.И.* Теория нейронных сетей: Учебное пособие для вузов. Кн.1. М.: ИПРЖР, 2000. 416 с.
8. *Kruglov I.A., Mishulina O.A., Bakirov M.B.* Quantile based decision making rule of the neural networks committee for ill-posed approximation problems // Neurocomputing. 2012. Vol. 96. P.74-82.

REFERENCES:

1. *Kulik S. D.* Informatsionnaya bezopasnost' AFIPS // Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskiye nauki. 2003. T. 33. № 4. P. 238.
2. *Kulik S. D.*Algoritmy raspoznavaniya obrazov i modelirovaniye avtomatizirovannykh faktograficheskikh informatsionno - poiskovykh sistem // Neurocomputers: razrabotka i primeneniye. 2007. № 2-3. P. 67-82.
3. *KulikS. D.*Primeneniye neyrosetevogo podkhoda v informatsionnykh i ekspertnykh sistemakh // Neurocomputers: raz-rabotka i primeneniye. 2007. № 2-3. P. 93-118.
4. *Wentzel E.S.* Teoriya veroyatnostey. M.: Vysshaya shkola. 2001. 575 p.
5. *Kolmogorov A.N.* Osnovnyye ponyatiya teorii veroyatnostey. M.: FAZIS, 1998. 144 p.
6. *Feller W.* An introduction to probability theory and its applications, Vol.1, 3nded., John Wiley & Sons. New York, 1968. xvii+509 pp.
7. *Galushkin A. I.* Teoriya neyronnykh setey: Uchebnoye posobiye dlya vuzov. Kn.1. M .: IPRZHR. 2000. 416 p.
8. *Kruglov I.A., Mishulina O.A., Bakirov M.B.* Quantile based decision making rule of the neural networks committee for ill-posed approximation problems // Neurocomputing. 2012. Vol. 96. P.74-82.