УДК 004.056.53:004.492:004.89

М.П. Комар

Тернопольский национальный экономический университет, Тернополь

МЕТОД ПОСТРОЕНИЯ СОВОКУПНОГО КЛАССИФИКАТОРА ТРАФИКА ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЕЙ ДЛЯ ИЕРАРХИЧЕСКОЙ КЛАССИФИКАЦИИ КОМПЬЮТЕРНЫХ АТАК

Предложен метод построения совокупного классификатора трафика информационно-телекоммуникационных сетей для иерархической классификации компьютерных атак. Данный метод базируется на предварительной обработке информации с использованием метода главных компонент и на объединении нейросетевых детекторов, каждый из которых обучен на определенный тип атаки. Представлены результаты экспериментальных исследований.

Ключевые слова: *классификатор трафика, информационно-телекоммуникационные сети, иерархическая классификация компьютерных атак, нейросетевой детектор, сжатие данных, метод главных компонент.*

Введение

В информационно-коммуникационных сетях (ИКС) защита информации становится еще более актуальной проблемой в связи с развитием и распространением глобальных вычислительных сетей, территориально распределенных информационных комплексов и систем с удаленным управлением доступом к информационным ресурсам. Весомым аргументом для повышения внимания к вопросам безопасности ИКС является бурное развитие программно-аппаратных методов и средств, способных скрыто существовать в системе и осуществлять потенциально любые несанкционированные действия (процессы), которые препятствуют нормальной работе пользователя и самой системы и непосредственно наносит вред свойствам информации (конфиденциальности, доступности, целостности). Несмотря на разработку специальных программно-аппаратных средств защиты от воздействия угроз на информационные ресурсы автоматизированных систем, количество новых методов реализации атак постоянно растет [1].

Сложившаяся ситуация стимулирует поиск и разработку новых подходов, направленных на повышение уровня защищенности ИКС от вредных воздействий. Методы искусственного интеллекта позволяют создавать принципиально новые средства обнаружения сетевых атак на ИКС, основанные на применении нейронных сетей [2, 3] и искусственных иммунных систем [4], а также строить на их основе интеллектуальные информационные технологии контроля и классификации трафика ИКС для обнаружения атак, что позволит повысить уровень защищенности ИКС от несанкционированного воздействия.

Сжатие входных данных на основе метода главных компонент

Как уже отмечалось в [5 – 7], выделяют 41 параметр сетевого соединения в ИКС, которые поступают на вход нейросетевого детектора для его обучения и классификации трафика, с целью выявления сетевых атак. Проведенные эксперименты показывают способность разработанной системы обнаруживать сетевые атаки. Однако, часть атак, таких как loadmodule, perl, spy, guess_passwd остаются практически не детектируемыми. Во многом это возникает по причине того, что используемая для теста база соединений KDD-99 [8] содержит недостаточное количество записей об этих атаках и данных, на которых можно обучить нейронную сеть.

Для того чтобы улучшить качество обучения нейронных сетей на «редких» типах атак, предлагается использовать метод главных компонент [9, 10] для сокращения размера данных при обучении и классификации сетевого трафика [11 – 13].

Метод главных компонент (principal component analysis, PCA) является одним из основных способов уменьшения размерность данных, потеряв при этом наименьшее количество информации. В настоящее время метод применяется для решения многих инженерных задач, таких как распознавание образов, компьютерное зрение, сжатие данных и т. д. Вычисление главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных. Метод главных компонент состоит в линейном ортогональном преобразовании входного вектора X размерности n в выходной вектор Y размерности p, где p<n [145]. Совокупность входных образов представим в виде матрицы:

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_n^1 \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \dots & \dots & \dots & \dots \\ x_1^L & x_2^L & \dots & x_n^L \end{bmatrix},$$

где $X^k = \left(x_1^k, x_2^k \dots x_n^k\right)$ соответствует k-му входному образу, L - общее количество образов.

Будем считать, что матрица X является центрированной, то есть вектор математических ожиданий $\mu=0$. Матрица ковариаций входных данных X определяется как

$$K = E\{XX^T\} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix},$$

где σ_{ij} - ковариация между i-ой и j-ой компонентой входных образов.

Метод главных компонент состоит в нахождении таких линейных комбинаций исходных переменных X, что выходные переменные Y (главные компоненты) некоррелированные между собой, упорядочены по возрастанию дисперсии и сумма дисперсий входных образов (в случае n=p) после преобразования остается без изменений. Тогда подмножество первых p главных компонент характеризует большую часть общей дисперсии. В результате получается сжатое представление входной информации.

В общем виде метод главных компонент можно представить в виде следующей формулы:

$$Y = XP,$$

где P – ортонормированная матрица линейного преобразования размерности n×p, столбцы которой соответствуют собственным векторам ковариационной матрицы. Такое преобразование можно представить в следующем виде:

$$KP = P\lambda$$

Или

$$P^T KP = \lambda$$

где P – диагональная матрица собственных значений ковариационной матрицы K.

Определим число главных компонент, которые будем использовать для анализа сетевого трафика. Для этого рассмотрим следующий критерий информативности [145]:

$$J = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{\lambda_1 + \lambda_2 + \dots + \lambda_n}$$

Анализируя при помощи этого выражения распределение количества информации, содержащейся в каждой последующей главной компоненте, можно определить число главных компонент p, которые целесообразно использовать для дальнейшего анализа без существенной потери информативности J.

На рис. 1 представлена зависимость количества информации от числа главных компонент.

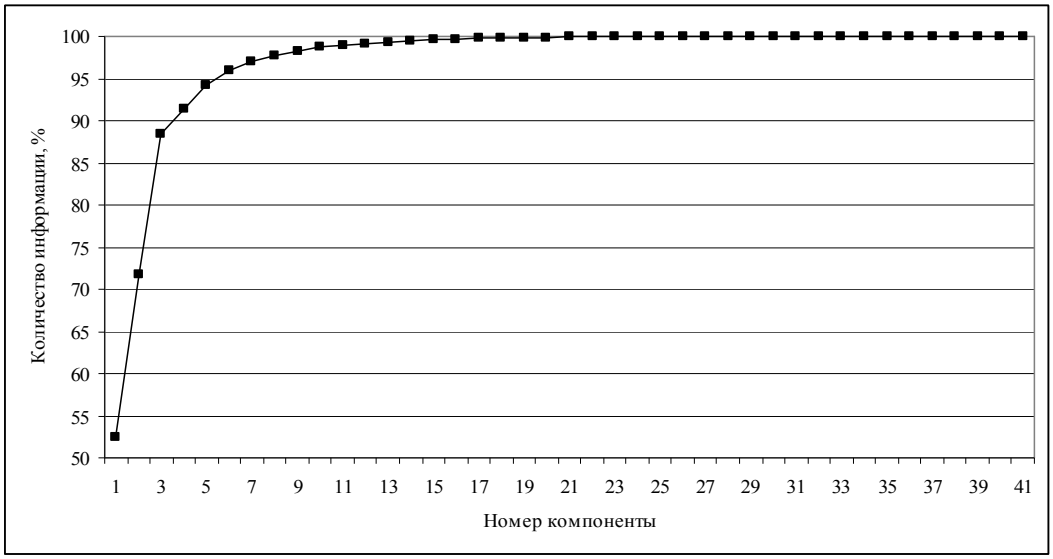

Рис. 1. Зависимость количества информации от числа главных компонент

Как видно из рис. 1 12 первых главных компонент содержат более 99% информации о сетевом трафике. В остальных 30 компонентах содержится меньше 1% информации, и из соображения целесообразности их можно исключить из анализа. Таким образом, для успешного анализа сетевого трафика достаточно использовать 12 первых главных компонент, в которых содержится более 99% информации о сетевом соединении, а не 41 параметр. Это позволяет существенно ускорить как процесс обучения нейросетевого детектора, так и процесс анализа сетевого трафика. Для этого, к выделенным из сетевого трафика данным применяем сначала метод главных компонент, а затем, подаем полученные данные на вход нейронной сети.

Совокупный классификатор для иерархической классификации компьютерных атак

Совокупным нейросетевым детектором называется такой детектор, который состоит из множества нейросетевых детекторов, каждый из которых обучен на определенном типе атак.

В [5, 6] представлен метод построения нейросетевых детекторов для обнаружения компьютерных атак.

Такой детектор состоит их трехслойной нейронной сети, где в качестве скрытого слоя используется слой Кохонена. Детектор обучается на обучающей выборке, состоящей из параметров сетевого трафика, причем 20% обучающей выборки составляют нормальные сетевые соединения, а 80% - один из типов атак.

Поскольку выделяют 22 разновидности сетевых атак, которые объединены в 4 класса (DoS, U2R, R2L и Probe) [8], то применяются 22 нейросетевых детектора, обученные на обнаружение каждой из 22 разновидностей сетевых атак. Схема функционирования совокупного нейросетевого детектора для обнаружения сетевых атак представлена в [5, 6].

Значения параметров сетевых соединений подаются на совокупный нейросетевой детектор. Активность одного из детекторов укажет на обнаружение аномального соединения, тип которого соответствует той сетевой атаке, на обнаружение которой обучался данный детектор. Если ни один из детекторов не реагирует на данное соединение, то данное соединение считается нормальным. Алгоритм функционирования детекторов обнаружения сетевых вторжений на основе совокупности нейросетевых детекторов представлен

На рис. 2 представлена схема построения совокупного классификатора для иерархической классификации компьютерных атак, который базируется на сжатии информации с использованием метода главных компонент и на объединении нейросетевых детекторов.

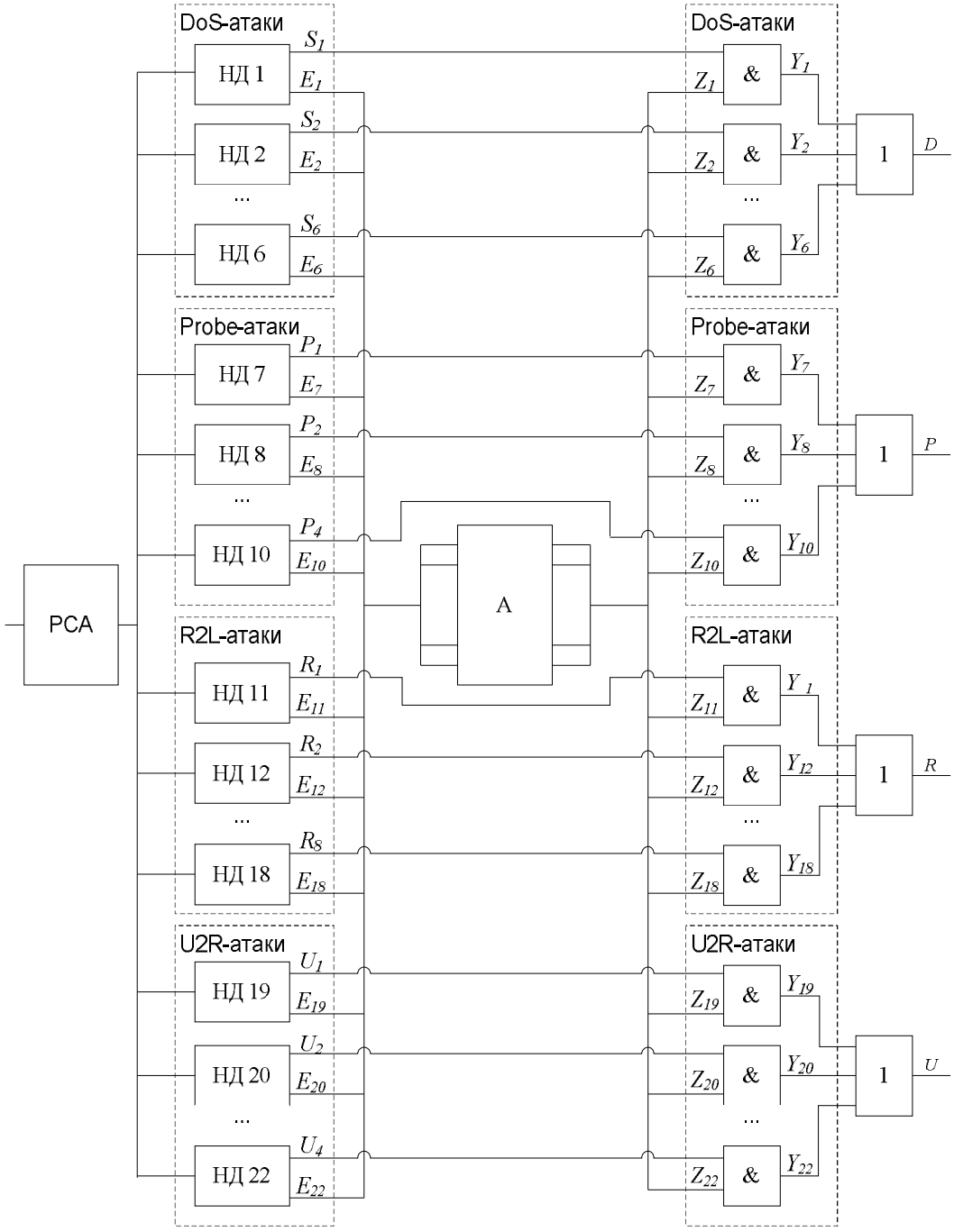

Рис. 2. Схема совокупного классификатора для иерархической классификации компьютерных атак

Сжатый набор входных данных размерностью 12 поступает на нейросетевые детекторы, каждый из которых обучен на соответствующий тип атак. В результате, если детектор обнаруживает атаку, то выходное значение его первого выхода устанавливается в единичное значение. Для устранения конфликтов в работе такого классификатора, когда несколько детекторов устанавливаются в единичное состояние, на второй выход каждого детектора передается минимальное евклидово расстояние между входным образом и весовыми векторами соответствующего детектора:

$$E_j = \min_j D_j =$$

$$= \min_i \sqrt{(x_1 - w_{1j})^2 + (x_2 - w_{2j})^2 + ... + (x_{12} - w_{12j})^2}.$$

Информация о минимальном евклидовом расстоянии поступает с каждого детектора на арбитр, который определяет детектор с номером k , имеющий минимальную евклидову метрику:

$$E_k = \min E_j,\ j = \overline{1,22}.$$

В результате k-й выход арбитра устанавливается в единичное состояние, а остальные выходы – в нулевое состояние:

$$Z_i = \begin{cases} 1, \text{если } i = k; \\ 0, \text{иначе.} \end{cases}$$

На выходах логических элементов «И» определяется тип атаки:

$$Y_i = F_i Z_i.$$

где $F_i = \begin{cases} S_i, \text{если } i = \overline{1,6}; \\ P_i, \text{если } i = \overline{7,10}; \\ R_i, \text{если } i = \overline{11,18}; \\ U_i, \text{если } i = \overline{19,22}. \end{cases}$

Выходы логических элементов «ИЛИ» определяют класс атаки:

$$D = \overset{6}{\underset{i=1}{\forall}} Y_i,\ P = \overset{10}{\underset{i=7}{\forall}} Y_i,\ R = \overset{18}{\underset{i=11}{\forall}} Y_i,\ U = \overset{22}{\underset{i=19}{\forall}} Y_i,$$

где D – DoS-атака; P– Probe-атака; R– R2L-атака; U– U2R-атака.

Таким образом, предложен иерархической классификатор сетевых атак, который состоит совокупности нейросетевых детекторов, каждый из которых обучен на определенный тип атаки и позволяет определять тип и класс сетевых атак.

Экспериментальные исследования построения совокупного классификатора с использованием метода главных компонент

Экспериментальные исследования проводились на основе методики, представленной в табл. 1.

При обучении нейросетевого детектора использовались следующие данные: 80% обучающей выборки составляют данные сетевой атаки, а 20% обучающей выборки составляют данные нормального соединения.

Таблица 1

Условия проведения экспериментов

Размер обучающей выборки	80 векторов (64 – атака, 16 – нормальное соединение)	
Адаптация параметров соединения	параметр protocol_type: tcp – 1; udp – 2; icmp – 3… параметр service: http – 1; smtp – 2; … ; telnet – 5 …	
Структура нейронной сети LVQ	12-10-2	
Размер тестовой выборки	DOS	391458
	Probe	4107
	R2L	1126
	U2R	52
	Normal	97278

Проведенные результаты показали, что при использовании метода главных компонент для предобработки данных о сетевом соединении, качество обнаружения значительно улучшилось.

На рис. 3 представлен сравнительный анализ обобщенных результатов обнаружения сетевых атак по классах с использованием метода главных компонент и без него.

Как видно из полученных результатов, качество обнаружения удалось значительно увеличить благодаря применению метода главных компонент к параметрам сетевого трафика ИКС.

Выводы

Таким образом, в статье предложен метод построения совокупного классификатора для иерархической классификации компьютерных атак, который базируется на предварительной обработке информации с использованием метода главных компонент и на объединении нейросетевых детекторов, каждый из которых обучен на определенный тип атаки. Применение метода главных компонент для сокращения размера данных для анализа сетевого трафика с целью выявления сетевых атак позволило повысить качество обнаружения сетевых атак на компьютерные системы, а также быстродействие системы за счет сокращения анализируемых данных.

Список литературы

1. Юдін О.К. Методи виявлення атак до інформаційних ресурсів автоматизованих систем / О.К. Юдін, Е.О. Коновалов, І.Є. Рогоза // Захист інформації. – 2010. – №2. – С. 5-12.

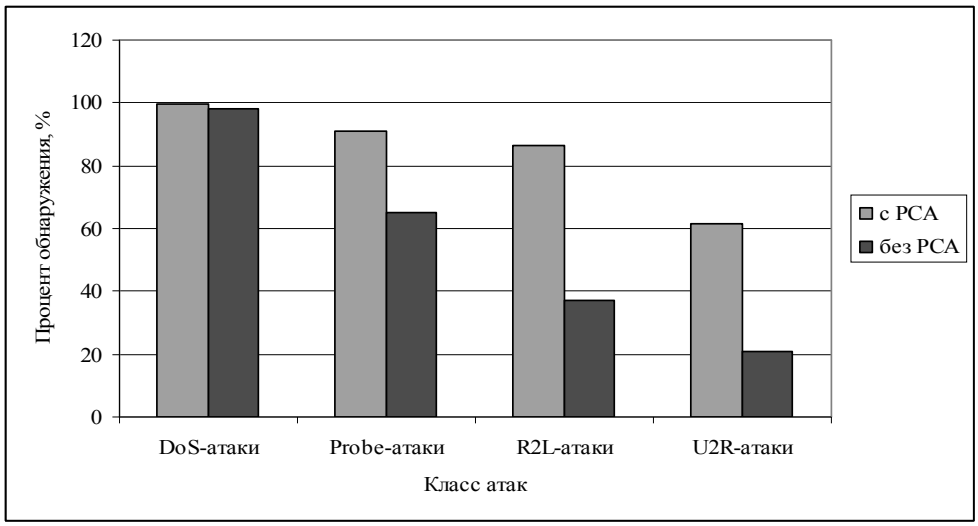

Рис. 3. Сравнительный анализ обобщенных результатов обнаружения сетевых атак по классах

2. Головко В.А. Нейронные сети: обучение, организация и применение : учеб. пособие / В. А. Головко. – М.: Радиотехника, 2001. – 256 с.

3. Хайкин С. Нейронные сети: полный курс; пер. с англ. / С. Хайкин. – М.: Вильямс, 2006. – 1104 с.

4. Дасгупта Д. Искусственные иммунные системы и их применение / Д. Дасгупта. – М.: Физматлит, 2006. – 344 с.

5. Комар М.П. Система анализа сетевого трафика для обнаружения компьютерных атак / М.П. Комар // Вестник Брестского ГТУ: (Серия: физика, математика и информатика). – 2010. – №5. – С. 14–16.

6. Комар М.П. Методы искусственных нейронных сетей для обнаружения сетевых вторжений / М. Комар // Сб. тез. седьмой межд. НТК "Интернет - Образование - Наука " (ИОН-2010) – Винница: Винницкий национальный технический университет, 2010. – С. 410–413.

7. Комар М.П. Интеллектуализированная информационная технология обнаружения компьютерных атак / М.П. Комар, Д.И. Боднар, А.А. Саченко // Измерительная и вычислительная техника в технологических процессах. – 2010. – № 2. – С. 133–137.

8. KDD Cup 1999 Data [Электронный ресурс]. – Режим доступа: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

9. Jolliffe I. Principal component analysis / I.T. Jolliffe. – Springer, 2010. – 516 p.

10. Shilpa Lakhina. Feature Reduction using Principal Component Analysis for Effective Anomaly–Based Intrusion Detection on NSL-KDD / Shilpa Lakhina, Sini Joseph, Bhupendra Verma // International Journal of Engineering Science and Technology. – 2010. – Vol. 2(6). – P. 1790–1799.

11. Комар М.П. Інтелектуальна система виявлення мережевих атак на інформаційні ресурси на основі методу головних компонент / М.П. Комар // Системи обробки інформації. – 2011. – № 8 (98). – С. 203-207.

12. Комар М.П. Нейросетевой подход к обнаружению компьютерных атак на информационные ресурсы / М.П. Комар, И.О. Палий, Р.П. Шевчук, Т.Б. Федысив // Інформатика та математичні методи в моделюванні. – 2011. – Т. 1, №2. – С. 156-163.

13. Komar M. Intelligent system for detection of networking intrusion / M. Komar, V. Golovko, A. Sachenko, S. Bezobrazov // Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS-2011). – Prague (Czech Republic), 2011. – V1. – P. 374-377.

Поступила в редколлегию 23.02.2012

Рецензент: д-р техн. наук, проф. Н.П. Карпинский, Тернопольский национальный технический университет имени И. Пулюя, Тернополь.

МЕТОД ПОБУДОВИ СУКУПНОГО КЛАСИФІКАТОРА ТРАФІКУ ІНФОРМАЦІЙНО-ТЕЛЕКОМУНІКАЦІЙНИХ МЕРЕЖ ДЛЯ ІЄРАРХІЧНОЇ КЛАСИФІКАЦІЇ КОМП'ЮТЕРНИХ АТАК

М.П. Комар

Запропоновано метод побудови сукупного класифікатора трафіку інформаційно-телекомунікаційних мереж для ієрархічної класифікації комп'ютерних атак. Даний метод базується на попередній обробці інформації з використанням методу головних компонент і на об'єднанні нейромережевих детекторів, кожен з яких навчений на певному типу атаки. Представлені результати експериментальних досліджень.

Ключові слова: *класифікатор трафіку, інформаційно-телекомунікаційні мережі, ієрархічна класифікація комп'ютерних атак, нейромережевий детектор, стиснення даних, метод головних компонент.*

METHOD OF CUMULATIVE TRAFFIC CLASSIFIER DEVELOPMENT FOR HIERARCHICAL CLASSIFICATION OF COMPUTER ATTACKS IN THE TELECOMMUNICATION NETWORKS

M.P. Komar

A method of cumulative traffic classifier development for hierarchical classification of computer attacks in the telecommunication networks was proposed. This method is based on preliminary information processing using the principal components analysis and on the neural network detectors association, each of them is trained on a certain type of attack. The results of experiments was presented.

Keywords: *traffic classificator, information and telecommunication networks, hierarchical classification of computer attacks, neural network detector, data compression, the principal components analysis.*