По нему изображение будет сильно испорчено при понижении яркости на 5%, но человеческий глаз этого не заметит. В то же время изображения «со снегом» – изменение цвета отдельных точек, полос, «муаром» - будут признаны хорошими. Можно оценить изменение качества изображения по максимальному отклонению: $d(x,y) = \max_{ij} \left| x_{ij} - y_{ij} \right|$.

Эта мера чувствительна к биению отдельных пикселов. Т.е. существенно изменить значение только одного пиксела, что практически будет незаметно для глаза, изображение по этому критерию испорчено.

Мера, которую сейчас используют на практике – пиковое отношение сигнал/ шум (peak signal-to-noise ratio, PSNR) [3]:

$$d(x,y) = 10 lg \frac{255^2 \cdot n^2}{\sum_{i=1,j=1}^{n,n} \left(x_{ij} - y_{ij} \right)^2}.$$

Данная мера аналогична среднеквадратичному отклонению, но пользоваться ей удобней из-за логарифмического масштаба шкалы.

Качество изображения оценим по пиковому отношению сигнал/ шум *PSNR*.

Поскольку первый вариант больше зависим от исходных данных, вычислим сначала по экспериментальным данным его усредненные характеристики (таблица 1). Единицы измерения можно не учитывать, т.к. нам необходимы относительные величины.

Таблица 1

Критерий сравнения	**Опыт 1**	**Опыт 2**	**Опыт 3**	**Средние значения**
Размер контейнера	256x256			-
Размер сообщения	64x64			-
Евклидово расстояние	64	128	192	-
PSNR	50,3483	73,0174	68,0488	63,8048
Время квантования	3,104	2,444	2,253	2,60033
Время обратного квантования	0,05	0,05	0,05	0,05

Таблица 2

	Опыт 1		**Опыт 2**	
Критерий сравнения	**VQ**	**LVQ**	**VQ**	**LVQ**
Размер контейнера	256x256		256x256	
Размер сообщения	64x64		192x192	
PSNR	63,8	18-46	12,0	23,0
Время квантования	2,6	2,6	55,1	24,9
Время обратного квантования	0,05	69,91	3,27	731,50

Теперь проведем сравнение методов первого и второго вариантов (таблица 2) по выбранным критериям - времени квантования, времени обратного квантования, пиковому отношению сигнал/ шум.

Из табл.2 видно, что метод LVQ имеет значительно большее время обратного квантования, т.е. на приемной стороне необходимо затратить значительное время на извлечение изображения. Однако на приемной стороне выигрыш в скорости квантования примерно в 2 раза дает метод LVQ при увеличении объема встроенного сообщения. При малых объемах сообщения скорости квантования примерно одинаковы. Качество восстановленного изображения при малых объемах сообщения оставляет желать лучшего, однако при больших объемах сообщения метод LVQ превосходит метод VQ по качеству изображения в 2 раза.

Таким образом, при больших объемах встраиваемых данных несомненными преимуществами обладает метод с использованием нейронной сети LVQ.

Выводы

На основании проведенных исследований можно выработать рекомендации при использовании методов векторного квантования в стеганографии:

- При небольших объемах внедряемых данных лучшие результаты дает метод VQ с использованием адаптивной кодовой книги, например, при внедрении цифровых водяных знаков. Для подтверждения авторских прав необходимо иметь кодовую книгу. Кодовая книга этого метода небольшая, вполне может храниться вместе с логотипом внедренного изображения. Если уровень секретности тайного канала связи позволяет передавать кодовую книгу и у получателя не располагает временем для извлечения сообщения, то метод VQ может быть использован и для организации тайного канала связи.
- При значительных объемах внедряемых данных целесообразно применить метод `LVQ с фиксированной кодовой книгой, например, для организации тайного канала связи. Кодовая книга создается заранее и хранится на приемной и передающей сторонах без передачи по каналу связи, что повышает уровень секретности; уменьшается также время встраивания. При этом надо учесть, что время извлечения сообщения на приемной стороне все же возрастает в сотни раз, т.е. получатель должен иметь временной ресурс, получая взамен более высокое качество изображения.

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. Larisa Gorbashko. A Text Data Hiding and Recovering without Host Image. – Proceeding of PRIP'2005, Minsk, 2005.
2. Mukherjee D., Chae J.J., Mitra S.K. A source and Channel Coding Approach to Data Hiding with Application to Hiding Speech in Video. Proceeding of IEEE ICIP'98, Chicago, Oct. 1998, Vol.1, pp.348-352.
3. Ватолин Д. И др. Методы сжатия данных.-М.: Диалог-МИФИ, 2003.
4. Уэлстид С. Фракталы и вейвлеты для сжатия изображений в действии. – М.: Триумф, 2003.
5. Грибунин В.А. Цифровая стеганография. – М.: Эксмо, 2002.
6. Горбашко Л.А. Метод скрытия изображений с применением векторного квантования.
7. Larisa Gorbashko, Vladimir Golovko. A Steganographic Method Using Learning Vector Quantization. –Proceedings of ICNNAI'2006, Brest, 2006.

Статья поступила в редакцию 21.12.2006

УДК 004.8.032.26

Кочурко П.А.

РАСПОЗНАВАНИЕ КЛАССОВ СЕТЕВЫХ АТАК: ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ РАЗЛИЧНЫХ АРХИТЕКТУР

1. Введение

Среди задач системы защиты информации, реализующей механизмы обнаружения атак [1, стр. 96], выделяют в том числе качественное распознавание известных и, зачастую,

***Кочурко Павел Анатольевич**, аспирант кафедры интеллектуальных информационных технологий БрГТУ.*
Беларусь, Брестский государственный технический университет, 224017, Беларусь, г. Брест, ул. Московская, 267.

неизвестных атак. Недостаточно просто обнаружить атаку – необходимо в первую очередь определить её тип, потому что от этого во многом зависит дальнейший алгоритм действий системы и персонала по защите информации и нейтрализации последствий атаки. Если для противодействия одному типу атак достаточно закрыть сетевой порт, то для другого может понадобиться активный ответ.

Технологии и методы атакующих постоянно развиваются, и стопроцентной защиты сетевого периметра не может гарантировать ни одна существующая система обнаружения атак. Поэтому методы обнаружения и распознавания атак также постоянно развиваются, и данная тематика развивается в работах различных исследователей. Для распознавания атак чаще всего применяются методы сигнатурного поиска, которые имеют один главный недостаток – недостаточная гибкость при обнаружении и распознавания модифицированных атак. Активно исследуются такие подходы, как кластеризация, нечёткая логика, деревья решений, методы опорных векторов и др. Искусственные нейронные сети (ИНС) имеют потенциал для решения большого количества проблем, охватываемых другими современными подходами к обнаружению атак. Изначально ИНС были заявлены в качестве альтернативы компонентам статистического анализа систем выявления аномалий. В дальнейшем предлагались подходы, использующие ИНС для обнаружения злоупотреблений и распознавания сетевых атак. В данной работе предлагаются подходы к распознаванию атак на основе нейросетевых детекторов различных архитектур.

Специфика основных подходов к обнаружению атак – обнаружения аномальной сетевой активности и обнаружения злоупотреблений – накладывает соответствующие ограничения на их нейросетевую реализацию. При этом подсистема анализа трафика может включать в себя любое количество ИНС и связей между ними. В случае, если используются несколько ИНС, должна быть разработана схема определения результата исходя из нескольких результатов, представленных каждой ИНС.

Для тестирования предложенных методов проведен ряд экспериментов. В качестве входных данных для обучения и тестирования детекторов использовалась база данных 1999 KDD Cup [2]. Она представляет собой информацию о TCP-соединениях реальной локальной вычислительной сети Air Force's Research Laboratory из Рима, штат Нью-Йорк, на основе которых были смоделированы две недели сетевого трафика, включавшего неизвестные и известные атаки. Каждое соединение описывается 41 параметром – основными параметрами (длительность, протоколы, …), параметрами данных (количество логинов, системных обращений, …) и статистическим (количество подключений к данному сервису за последнее временное окно, …). Все соединения в базе данных подразделяются на пять классов: нормальные соединения; DOS-атаки (отказ в обслуживании); probe-атаки (сканирование портов и др.); U2R-атаки (неавторизованное получение привилегий root на данной системе); R2L-атаки (неавторизованный доступ с удалённой системы). Всего – 22 типа атак и нормальные соединения.

Статья организована следующим образом. В разделах 2-4 рассматриваются подходы к распознаванию атак на основе детекторов различных архитектур. В разделе 5 производится сравнение результатов с результатами аналогичных исследований и делаются выводы по дальнейшим направлениям исследования.

2. Детектор на основе LVQ

В упрощенном виде схему работы системы обнаружения сетевых атак можно представить в следующем виде. Производится перехват сетевого трафика; производится предварительная обработка трафика с выделением параметров TCP-соединений, которые поступают на обработку детекторами атак. Детекторов может быть много, и соединение может поступать на один из них в зависимости, например, от службы, указанной в параметрах соединения, а может – на все сразу, и решение будет приниматься консолидированно.

Векторный квантователь может быть использован в качестве детектора злоупотреблений для распознавания типа атак. Так как база данных KDD содержит атаки 22 типов плюс нормальные соединения, то получаем 23 класса, принадлежность к которым входных векторов можно определить. На вход LVQ поступает 41 параметр – по количеству параметров в записях KDD.

LVQ описанной выше архитектуры с 50 подклассами обучался по методу контролируемого конкурентного обучения на наборе данных, содержащем по 30 соединений всех типов для служб HTTP, PRIVATE и для всех служб вместе (ALL). Тестировался на всех соединениях служб HTTP и PRIVATE в десятипроцентной выборке KDD, а для ALL - по 100 векторов каждой атаки каждой службы из десятипроцентной выборки. Результаты показаны в таблицах 1-2.

Как видим, значения ошибок для детектора злоупотреблений на основе LVQ слишком высоки, а значит необходимо улучшение данного подхода.

Как указано выше, все соединения, представленные в базе данных KDD, принадлежат к одному из 23 типов, каждый из которых относится к одному из пяти классов – DOS, PROBE, R2L, U2R, NORMAL. Схему распознавания классов, аналогичную описанной схеме распознавания типов, можно реализовать по-разному:

- если в архитектуре LVQ на выходе анализировать не тип (23 выхода), а класс (5 выходов);
- так как каждый тип относится к одному из классов, то можно кроме типа сообщать ещё и класс результата.

Выведем общий результат распознавания атак различных классов при помощи детектора на основе LVQ (таблица 3).

Отметим, что данная технология может быть улучшена с использованием предварительной обработки данных с помощью PCA или nPCA-сетей.

Таблица 1. *Результаты тестирования качества распознавания некоторых типов атак с помощью LVQ (DR – detection rate)*

	ALL	HTTP	PRIVATE
	DR, %	DR, %	DR, %
back	0,00	99,59	–
ipsweep	80,32	100	79,41
neptune	81,83	87,50	80,83
nmap	0,00	–	79,83
phf	0,00	100	–
portsweep	89,71	100	96,83
satan	86,35	100	90,13
teardrop	0,00	–	99,89
normal	78,92	91,16	99,99

Таблица 2. *Общие результаты тестирования LVQ в качестве детектора злоупотреблений (FN – false negative, FP – false positive, MC – misclassification)*

	FN, %	FP, %	MC, %	Качество распознавания			
				dos, %	probe, %	r2l, %	u2r, %
ALL	8,43	21,08	29,02	72,28	79,19	45,88	37,82
HTTP	0,30	8,84	1,37	98,62	100,00	100,00	–
PRIVATE	0,01	0,001	17,60	81,02	92,45	–	–

3. Детектор на основе MLP

Для реализации технологии распознавания атак обучим многослойный персептрон из трёх слоёв с нелинейными функциями активации нейронов в скрытом (гиперболический тангенс) и выходном слое (сигмоидная). Количество нейрон-

ных элементов в распределительном слое соответствует количеству параметров в базе данных KDD, в выходном слое – количеству типов соединений. В процессе функционирования при подаче параметров соединения на вход сети на выходе наибольшее значение будет иметь нейрон, соответствующий типу данного соединения.

Для обучения и тестирования используем те же наборы данных, что и для LVQ. Для обучения используем метод обратного распространения ошибки с адаптивным шагом обучения и модификацией весовых коэффициентов, используя для выхода из локальных минимумов известный в теории оптимизации как метод "тяжелого шарика".

Результаты тестирования детекторов на основе MLP приведены в таблицах 3-4. Как видим, качество распознавания и обнаружения атак по сравнению с предыдущим подходом значительно выше.

4. Совокупный детектор на основе РНС

Рециркуляционные нейронные сети отличаются от других искусственных нейронных сетей тем, что информация, подающаяся на вход, в том же виде восстанавливается на выходе. Хорошие результаты показали нелинейные РНС в качестве детектора аномалий [3,4]: обучение РНС производится на нормальных соединениях таким образом, чтобы входные векторы на выходе восстанавливались в себя, при этом - чем соединение более похоже на нормальное, тем меньше ошибка реконструкции:

$$E^k = \sum_j \left(\bar{X}_j^k - X_j^k \right)^2, \qquad (1)$$

где X_j^k – j-й элемент k-го входного вектора, $\bar{X}_j^k$ – j-й элемент k-го выходного вектора. Если $E^k > T$, где T – некий заданный для данной РНС порог, то соединение признаётся аномалией, или атакой, иначе – нормальным соединением.

Одна РНС может применяться для определения принадлежности входного вектора к одному из двух классов – тому, на котором обучалась (класс A), или ко второму (класс $\bar{A}$), которому соответствуют далеко отстающие вектора:

$$\begin{cases} X^k \in A, \ \textit{если} \ \ E^k \le T, \\ X^k \in \bar{A}, \ \textit{если} \ \ E^k > T. \end{cases} \qquad (2)$$

Таблица 3. *Результаты тестирования качества распознавания некоторых типов атак с помощью MLP*

	ALL	HTTP	PRIVATE
	DR, %	DR, %	DR, %
back	99,00	99,68	-
ipsweep	97,37	100,00	100,00
neptune	99,98	100,00	99,90
nmap	91,43	-	100,00
phf	100,00	100,00	-
portsweep	97,11	100,00	98,90
satan	97,52	100,00	99,68
teardrop	0,00	-	99,90
normal	97,48	-	99,78

Таблица 4. *Общие результаты тестирования MLP в качестве детектора злоупотреблений*

	FN, %	FP, %	MC, %	Качество распознавания			
				dos, %	probe, %	r2l, %	u2r, %
ALL	1,34	2,52	4,38	97,95	96,81	81,44	94,45
HTTP	0,29	0,21	0,29	99,71	100,00	100,00	–
PRIVATE	0,00	0,01	0,12	99,90	99,27	–	–

Доказано [5, 6], что лучшие результаты при классификации (даже вопрос – «атака или нет?» - есть ни что иное, как определение принадлежности к классу атак или классу нормальных соединений; не говоря уже об определении класса атаки) дают независимые друг от друга классификаторы.

Основной проблемой в разработке систем из нескольких независимых детекторов или классификаторов становится вопрос выбора наиболее правдоподобного значения среди результатов, выдаваемых разными классификаторами (динамический выбор классификатора). В случае применения «слишком независимых» детекторов есть опасность, что построение общей оценки будет затруднено из-за несоизмеримости или несравнимости выходов детекторов. Значительно больше возможностей для построения совокупной оценки общего классификатора при использовании независимых детекторов одинаковой природы. В этом случае выходы каждого отдельного детектора сравнимы между собой, и могут применяться различные методы динамического выбора классификатора: средняя оценка, максимальный голос, метод выбора “a posteriori” и др. [5].

Общий классификатор состоит из N=5 частных детекторов, каждый из которых имеет порог T_i. Так как значения порогов на этапе обучения каждого из детекторов выбирались исходя из минимизации средней стоимости ошибки, то для приведения оценок детекторов к сравнимым значениям достаточно отмасштабировать ошибку реконструкции по порогу [9]. Тогда (2) запишется как:

$$\begin{cases} X^k \in A_i, \ \textit{если} \ \ \delta_i^k \le 1, \\ X^k \in \bar{A}_i, \ \textit{если} \ \ \delta_i^k > 1, \end{cases} \qquad (3)$$

где $\delta_i^k = \dfrac{E_i^k}{T_i}$ - относительная ошибка реконструкции. При этом, чем меньше δ_i^k, тем более вероятна принадлежность входного образа X^k к классу A_i. Поэтому можно выделить метод определения совокупной оценки – по минимальной относительной ошибке реконструкции:

$$\begin{cases} X^k \in A_m, \\ \delta_m^k = \min\limits_i \delta_i^k. \end{cases} \qquad (4)$$

Обучены частные детекторы для каждого из пяти классов атак. РНС обучались таким образом, чтобы восстанавливать соединения своего класса в аналогичные соединения с минимальной ошибкой реконструкции.

Результаты тестирования совокупного классификатора приведены в таблице 5.

Таблица 5. *Результаты обнаружения и распознавания атак совокупным классификатором*

	FP, %	FN, %	Качество распознавания			
			dos, %	probe, %	r2l, %	u2r, %
ALL	10,80	2,34	98,17	96,55	91,88	100
HTTP	0	0,08	99,75	100	100	–
FTP_DATA	0,66	1,09	100	100	96,66	100
TELNET	0	5,26	98,75	100	97,33	85,50

5. Сравнение результатов

Сравнивая результаты (рисунок 1), показываемые различными технологиями, отметим, что самый высокий уровень распознавания и классификации атак показывает технология совокупного классификатора на основе РНС, правда при достаточно высоком уровне FN. В свою очередь, очень хорошие результаты по уровню FN и FP показывают подходы на основе MLP, причем и качество классификации атак данными методами только чуть хуже, чем у совокупного классификатора.

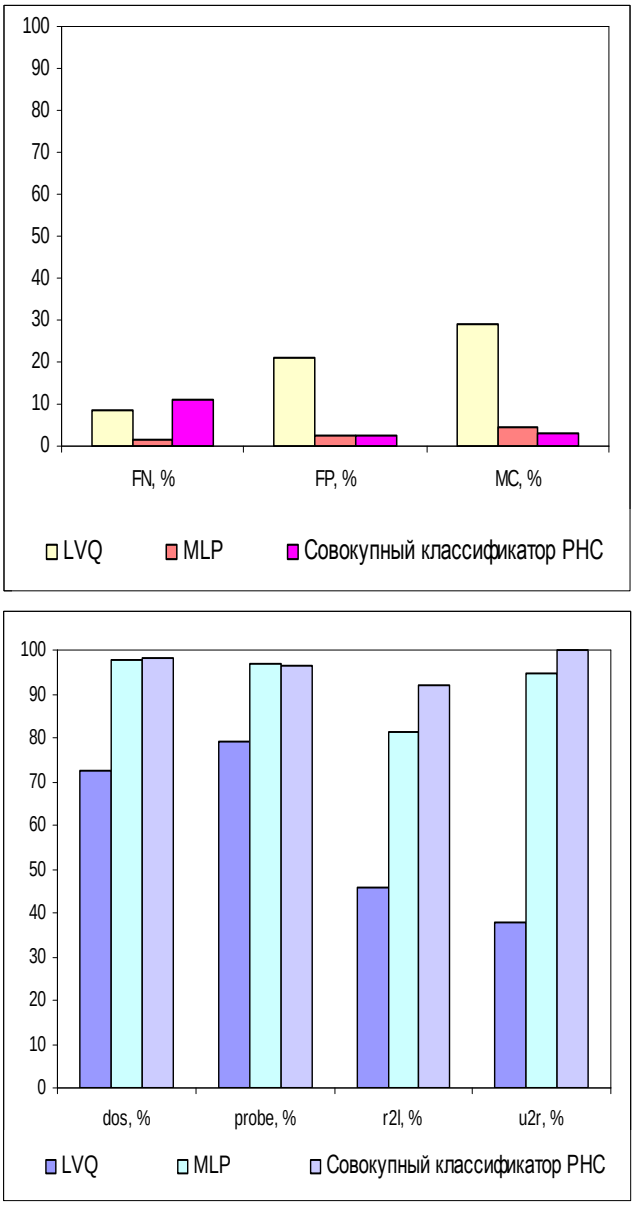

Рис. 1. *Сравнение результатов, показанных детекторами с различными нейросетями*

Если рассмотреть результаты с применением других технологий в других исследованиях (таблица 6), то видно, что качество классификации с помощью совокупного классификатора на основе РНС существенно улучшает все применяемые технологии. Особенно это заметно на атаках классов u2r и r2l.

Таблица 6. *Результаты распознавания классов атак в некоторых исследованиях* [7]

	dos, %	probe, %	r2l, %	u2r, %
MLP	97,2	88,7	5,6	73,2
Гауссовский классификатор	82,4	90,2	9,6	22,8
K-NN	97,3	87,6	6,4	29,8
Алгоритм ближайшего кластера	97,1	88,8	3,4	2,2
RBF	73,0	93,2	5,9	6,1
Лидер-алгоритм	97,2	83,8	1,0	6,6
Алгоритм гиперсферы	97,2	84,8	1,0	8,3
Fuzzy Art Map	97,0	77,2	3,7	6,1
Дерево решений C4.5	97,0	80,8	4,6	1,8
Победитель KDD-99 [8]	97,1	83,3	13,2	8,4

Можно сделать вывод, что представленные технологии могут с успехом применяться для распознавания атак. Предложенный метод распознавания классов атак совокупным классификатором на основе РНС показывает самые лучшие результаты в распознавании классов атак. Он является гибким, масштабируемым: одна и та же архитектура классификатора может применяться для классификации большего класса входных данных всего лишь добавлением дополнительного детектора. Данный метод может с успехом применяться для решения и других задач классификации и распознавания образов.

Исследования проводятся при поддержке Министерства образования Республики Беларусь и БРФФИ при НАН Беларуси.

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. Лукацкий, А. В. Обнаружение атак / А. В. Лукацкий. – СПб.: БХВ-Петербург, 2003. – 596 с.
2. KDD Cup 1999 Data / The UCI KDD Archive, Information and Computer Science. - University of California, Irvine, 1999.
3. Некоторые аспекты применения нейронных сетей для обнаружения сетевых атак / В. А. Головко [и др.] // Вестник Брестского государственного технического университета. – 2004. – №5 (29): Физика, математика, информатика. - с. 35-39.
4. Кочурко, П. А. Нейросетевой детектор аномалий / П. А. Кочурко // Известия Белорусской инженерной академии – 2005. – № 1(19)/2. – с. 78-81.
5. Selection of image classifier / G. Giacinto [et al.] // Electron. – 2000. – №26(5). – pp. 420-422.
6. Methods for combining multiple classifiers and their applications to handwriting recognition / L. Xu [et al.] // IEEE Trans. Syst. Man Cybernitics. – 1992. – №22. – pp. 418-435.
7. Sabhnani, M. Application of Machine Learning Algorithms to KDD Intrusion detection dataset within Misuse detection context / M.Sabhnani, G. Serpen // The international conference on Machine Learning: Models, technologies and Applications: proceedings, 2003. – 2003. - pp. 209-215.
8. Pfahringer, B. Winning the KDD99 Classification Cup: Bagged Boosting / B. Pfahringer // SIGKDD Explorations. – 2000. – Vol. 1, №2. - pp. 65-66.
9. Кочурко, П. А. Совокупность детекторов на основе рециркуляционных нейронных сетей для распознавания класса сетевых атак / П. А. Кочурко // Вестник Брестского государственного технического университета. – 2005. – №5 (35): Физика, математика, информатика. - с. 61-66.

Статья поступила в редакцию 21.12.2006