Ю. Г. Емельянова, А. А. Талалаев, И. П. Тищенко, В. П. Фраленко, Нейросетевая технология обнаружения сетевых атак на информационные ресурсы, *Программные системы: теория и приложения*, 2011, том 2, выпуск 3, 3–15

ISSN 2079-3316 ПРОГРАММНЫЕ СИСТЕМЫ: ТЕОРИЯ И ПРИЛОЖЕНИЯ № 3(7), 2011, с. 3–15

УДК 004.042

Ю. Г. Емельянова, А. А. Талалаев, И. П. Тищенко,
В. П. Фраленко

Нейросетевая технология обнаружения сетевых атак на информационные ресурсы

Аннотация. Статья описывает нейросетевой подход, сочетающий быстроту обработки сетевого трафика за счет сжатия признаков и высокую точность классификации сетевых атак. Приведены результаты экспериментальных исследований.

Ключевые слова и фразы: искусственная нейронная сеть, сетевая атака, вычислительные системы, выделение признаков, классификация.

Введение

Обнаружение сетевых атак связано с выделением большого числа признаков, по которым можно проводить классификацию. Так, например, в общедоступной базе KDD99, содержащей порядка 5 миллионов классифицированных по 22 типам экземпляров атак (классов), используется 41 признак [1]. Все признаки информационно неравнозначны, причем уточнить их истинную значимость можно только после проведения дополнительных исследований. Задача оптимизации числа признаков является неотъемлемой частью процесса распознавания. В реальных условиях система должна автоматически выделять признаки и использовать их для решения задачи обнаружения атаки и определения ее типа. В настоящее время разрабатывается большое количество различных технологий защиты компьютерных сетей, основанных на искусственных нейронных сетях (ИНС) и статистическом анализе. К их недостаткам можно отнести уязвимость к новым атакам, низкую точность и скорость работы. В настоящей работе основной упор делается на сжатие пространства признаков для последующего использования в нейросетевой системе обнаружения атак. В качестве аппарата для сжатия используются метод главных компонент (МГК) и рециркуляционная нейронная сеть (РНС).

1. Используемые подходы к сжатию признаков

Один из подходов связан с применением МГК [2]. Метод позволяет преобразовывать систему координат для обеспечения максимального разделения объектов на классы. Второе предназначение метода — это сжатие пространства признаков. Пусть имеется n признаков, которые описывают объекты из множества P, d_α — расстояние от объекта $P_\alpha \in P$ до прямой AB, отвечающей требованию $\sum_{\alpha=1}^{|P|} {d_\alpha}^2 \longrightarrow min$.

Применим следующую процедуру:

- составим матрицу ковариаций C размерности $n \times n$: $c_{ij} = \frac{1}{|P|-1} \sum_{\alpha=1}^{|P|} ({x_i}^\alpha - \overline{x_i})({x_j}^\alpha - \overline{x_j})$, $i, j = 1, \ldots, n$, причем $c_{ij} = c_{ji}$, $\overline{x_m}$ — среднее значение параметра x_m;
- найдем собственные числа матрицы, решая характеристическое уравнение $det(C - \lambda E) = 0$, где λ — вектор собственных чисел матрицы C, E — единичная матрица;
- определим собственные вектора из равенства $CX = \lambda X$;
- введем новую систему ортогональных векторов, развернутую в пространстве относительно исходной системы координат. Проведем нормализацию для получения ортонормированных векторов, представляющих новую систему координат, и пересчитаем значения признаков.

Главными компонентами являются собственные числа λ, которые пропорциональны величинам дисперсии признаков. Пространство признаков можно сжать путем отбора k собственных чисел, имеющих наибольшее значение.

Второй подход основан на применении РНС с двумя слоями нейронов [3]. Первый слой, состоящий из k нейронов, позволяет управлять числом информационных признаков, а второй содержит n нейронов и производит фильтрацию данных (рис. 1). Обучение можно производить любым доступным способом, например, алгоритмом обратного распространения ошибки.

Для всего множества объектов P создается единственная нейронная сеть, на которую в случайном порядке подаются описания объектов этого множества. Обучение происходит до тех пор, пока для каждого P_α не будет получена его приближенная копия на выходном слое нейронов. Настройки первого слоя нейронов позволяют получать сжатую до k признаков форму представления любого входного

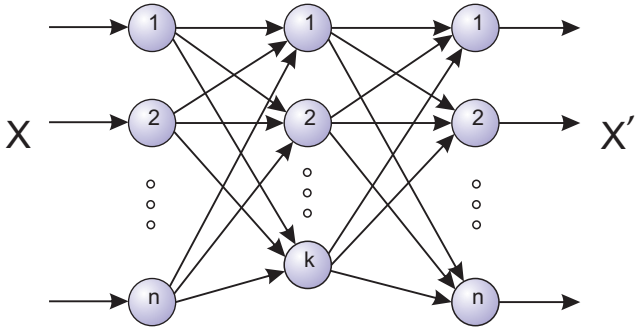

Рис. 1. Рециркуляционная нейронная сеть

n-мерного объекта. Эти признаки в дальнейшем будем назвать главными компонентами, поскольку РНС является нейросетевым аналогом МГК. Этот факт подтверждают и проведенные эксперименты.

2. Задача фильтрации сетевых атак

Задача заключается в разработке нейросетевого метода и программного обеспечения для обнаружения вторжений на основе нейросетевой технологии мониторинга аномальной сетевой активности.

Обнаружить атаку можно двумя методами [4]:

- *Сигнатурный метод* сводится к поиску признаков уже известных атак. Преимущество сигнатурного метода в том, что он практически не подвержен ложным срабатываниям. Недостатком этого метода является невозможность обнаруживать незаложенные в систему атаки.
- *Метод поиска аномалий* позволяет реагировать на ранее неизвестные атаки, но подвержен ложным срабатываниям и требует точной настройки для каждого наблюдаемого объекта.

Предлагается свести задачу обучения нейросетевой системы и отражения атак к следующей схеме (рис. 2):

(1) поиск и извлечение информативных признаков,
(2) сжатие признаков с помощью МГК или РНС,
(3) обучение двухслойного персептрона и сети Кохонена на базе выделенных информационных векторов признаков,
(4) включение в систему и работа в режиме отражения атак:

а. сжатие выбранных параметров с помощью МГК или РНС,
б. классификация нового информационного объекта.

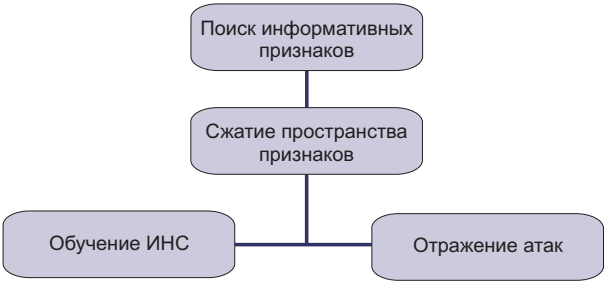

Рис. 2. Схема обучения и работы системы

Каждая запись в базе KDD99 представляет собой образ сетевого соединения. Соединение — последовательность TCP пакетов за некоторое конечное время, моменты начала и завершения которого четко определены, в течение которого данные передаются от IP-адреса источника на IP-адрес приемника, используя некоторый определенный протокол. Отдельная запись KDD99 включает 41 информационный признак и промаркирована как «атака» или «не атака». Например, первый параметр определяет длительность соединения, второй – – указывает используемый протокол, третий — целевую службу и т.д. При этом атаки делятся на четыре основные категории: DoS, U2R, R2L и Probe:

- DoS — отказ в обслуживании, характеризуется генерацией большого объема трафика, что приводит к перегрузке и блокированию сервера;
- U2R предполагает получение зарегистрированным пользователем привилегий локального суперпользователя (администратора);
- R2L характеризуется получением доступа незарегистрированного пользователя к компьютеру со стороны удаленной машины;
- Probe заключается в сканировании портов с целью получения конфиденциальной информации.

Все записи, хранящиеся в базе, были разделены на два примерно равных по мощности непересекающихся подмножества — данные из

первого использовались для обучения ИНС, данные из второго подавались на распознавание. Выборка эталонов очень неравномерна (смотри табл. 1), что существенно ухудшает обучение разрабатываемой системы обнаружения атак. Лишь 10 из 22 классов атак обладают достаточным количеством эталонов.

Таблица 1. Выборка сетевых пакетов

Группа	Класс	Количество
normal	normal	972781
u2r	buffer_overflow	30
u2r	loadmodule	9
u2r	perl	3
dos	neptune	1072017
dos	smurf	2807886
r2l	guess_passwd	53
dos	Pod	264
dos	teardrop	979
probe	portsweep	10413
probe	ipsweep	12481
dos	land	21
r2l	ftp_write	8
dos	back	2203
r2l	imap	12
probe	satan	15892
r2l	phf	4
probe	nmap	2316
r2l	multihop	7
r2l	warezmaster	20
r2l	warezclient	1020
r2l	spy	2
u2r	rootkit	10

3. Обобщенный подход

Малый объем обучающей выборки для некоторых классов атак (табл. 1) обуславливает целесообразность проведения поэтапной обработки данных. В ходе исследований был проведен ряд экспериментов, суть которых заключалась в определении самого факта атаки, без определения конкретной группы атаки или ее класса. Эксперименты проводились с использованием различных подходов, таких как: сети Кохонена и сети прямого распространения. Результаты данной классификации приведены в табл. 2. Тестирование показывает достаточно высокие показатели качества распознавания.

Таблица 2. Точность и полнота определения сетевых атак

Метод	Класс	Полнота	Точность
Сеть прямого распространения	норма	0.9994	0.9871
	атака	0.9946	0.9997
Сеть Кохонена	норма	0.9796	0.9884
	атака	0.9953	0.9917

4. Тестирование двухслойного персептрона

Результаты, полученные в ходе тестирования двухслойного персептрона (содержит сто нейронов в первом слое и три во втором), представлены в табл. 3. Всего 1.22% пакетов класса «normal» были отнесены ошибочно к той или иной атаке. Относительно высокая точность определения типа атак при достаточной полноте была достигнута на атаках класса «neptune», «smurf» и «land», чуть хуже определяются атаки класса «teardrop», «portsweep» и «ipsweep».

С использованием персептрона были проведены дополнительные эксперименты по классификации входного вектора по группам атак (табл. 4–5).

5. Тестирование нейронной сети Кохонена

Эксперимент был проведен также с использованием сети Кохонена, результаты тестирования приведены в табл. 6. Сеть показала результаты аналогичные двухслойному персептрону. Распознавание атак типа perl и phf затруднено, причиной этого является очень малый объем обучающей выборки для данных классов.

Таблица 3. Точность и полнота определения сетевых атак по классам (двухслойный персептрон, 100 нейронов)

Группа	Класс	Полнота	Точность
normal	normal	0.9878	0.9998
u2r	buffer_overflow	1.0000	0.0458
u2r	loadmodule	1.0000	0.0208
u2r	perl	1.0000	0.2857
dos	neptune	0.9999	1.0000
dos	smurf	0.9999	1.0000
r2l	guess_passwd	1.0000	0.7067
dos	pod	1.0000	0.5514
dos	teardrop	1.0000	0.9975
probe	portsweep	1.0000	0.9341
probe	ipsweep	0.9955	0.9271
dos	land	1.0000	0.8947
r2l	ftp_write	1.0000	0.0396
dos	back	0.9990	0.8743
r2l	imap	1.0000	0.1395
probe	satan	0.9977	0.9404
r2l	phf	1.0000	0.0236
probe	nmap	1.0000	0.7694
r2l	multihop	1.0000	0.2500
r2l	warezmaster	0.9500	0.0884
r2l	warezclient	0.9392	0.2324
r2l	spy	1.0000	0.0260
u2r	rootkit	1.0000	0.0063

6. Эксперименты с использованием расстояния Евклида–Махаланобиса

В качестве альтернативного нейронным сетям метода классификации было проведено тестирование с использованием расстояния Евклида–Махаланобиса. Результаты тестирования продемонстрированы в табл. 7.

Таблица 4. Классификация по группам атак (двухслойный персептрон, 100 нейронов)

Группа	Полнота	Точность
normal	0.9930	0.9968
dos	0.9985	1.0000
u2r	0.9655	0.0294
r2l	0.9707	0.2576
probe	0.9991	0.9669

Таблица 5. Уточнение классификации выявленной ранее атаки (двухслойный персептрон, 100 нейронов)

Группа	Полнота	Точность
dos	0.9999	1.0000
u2r	1.0000	0.3625
r2l	0.9769	0.9573
probe	0.9998	0.9992

7. Эксперименты на репрезентативной выборке

В данном эксперименте из множества классов атак были удалены классы, обладающие малым количеством прецедентов. Результаты тестирования отражены в табл. 8.

Полученные результаты подтверждают, что качество классификации напрямую зависит от количества эталонов отдельных классов в обучающей выборке. При малом числе эталонов ошибки возникают, в том числе, и с высокой уверенностью классификатора.

8. Архитектура модуля нейросетевого мониторинга сетевых атак

Представленная нейросетевая система внедрена в состав пакета IDS Snort [5]. Snort — это сетевая система обнаружения вторжений (IDS), способная выполнять в режиме реального времени анализ трафика, передаваемого по контролируемому интерфейсу, с целью обнаружения попыток взлома или попыток поиска уязвимостей (таких,

Таблица 6. Точность и полнота определения сетевых атак по классам (сеть Кохонена)

Группа	Класс	Полнота	Точность
normal	normal	0.9966	0.9969
u2r	buffer_overflow	0.7500	0.0325
u2r	loadmodule	0.5000	1.0000
u2r	perl	0.5000	1.0000
dos	neptune	0.9999	1.0000
dos	smurf	0.9992	1.0000
r2l	guess_passwd	0.9623	0.1683
dos	pod	0.7843	0.2492
dos	teardrop	0.9975	1.0000
probe	portsweep	0.9842	0.9987
probe	ipsweep	0.9704	0.9920
dos	land	1.0000	1.0000
r2l	ftp_write	0.2500	1.0000
dos	back	0.6633	0.7224
r2l	imap	0.8333	0.3030
probe	satan	0.9895	0.9795
r2l	phf	1.0000	0.0240
probe	nmap	0.9253	0.6871
r2l	multihop	0.5714	1.0000
r2l	warezmaster	0.9500	1.0000
r2l	warezclient	0.6912	0.8226
r2l	spy	0.5000	1.0000
u2r	rootkit	0.1429	0.1667

как переполнение буфера, сканирование портов, CGI-атаки, идентификация операционной системы, идентификация версий используемых сетевых сервисов и др.).

Созданный модуль мониторинга следует рассматривать как модуль эвристического анализа. Система правил, задаваемая пользователем Snort, в отличие от искусственных нейронных сетей, имеет значительно меньшую вероятность ложных срабатываний, однако не позволяет выявлять аномальную активность, не зафиксированную в

Таблица 7. Точность и полнота определения сетевых атак по классам (расстояние Евклида–Махаланобиса)

Группа	Класс	Полнота	Точность
normal	normal	0.9686	0.9946
u2r	buffer_overflow	0.9375	0.1546
u2r	loadmodule	1.0000	0.7500
u2r	perl	1.0000	1.0000
dos	neptune	0.9705	0.9996
dos	smurf	0.9986	0.9999
r2l	guess_passwd	0.9615	1.0000
dos	pod	0.9771	0.9275
dos	teardrop	0.9898	0.9939
probe	portsweep	0.9185	0.3441
probe	ipsweep	0.9424	0.9192
dos	land	1.0000	0.6000
r2l	ftp_write	1.0000	0.8000
dos	back	0.9510	1.0000
r2l	imap	1.0000	1.0000
probe	satan	0.9138	0.5096
r2l	phf	1.0000	1.0000
probe	nmap	0.4465	0.0548
r2l	multihop	1.0000	1.0000
r2l	warezmaster	1.0000	0.9167
r2l	warezclient	0.9471	0.0678
r2l	spy	1.0000	1.0000
u2r	rootkit	1.0000	0.0746

базе правил в виде точных сигнатур. Исходя из этого, можно отметить, что методы обнаружения атак на основе правил и ИНС взаимно дополняют друг друга.

Общий механизм выявления сетевых атак с использованием модуля Neuronet выглядит следующим образом:

(1) система Snort перехватывает пакет;
(2) по группе правил системы Snort происходит вызов модуля для определения характера пришедшего пакета;

Таблица 8. Точность и полнота определения сетевых атак (репрезентативная выборка)

Группа	Класс	Сеть прямого распространения		Кохонен		Расстояние Евклида–Махаланобиса	
		Полнота	Точность	Полнота	Точность	Полнота	Точность
normal	normal	0.9896	0.9991	0.9961	0.9958	0.9689	0.9946
dos	neptune	0.9999	1.0000	0.9999	1.0000	0.9705	0.9996
dos	smurf	0.9997	1.0000	0.9994	1.0000	0.9986	0.9999
dos	pod	1.0000	0.7308	0.9549	0.0992	0.9771	0.9275
dos	teardrop	1.0000	0.9919	0.9959	0.8484	0.9918	0.9939
probe	portsweep	0.9985	0.9763	0.9862	0.9902	0.9185	0.3441
probe	ipsweep	0.9930	0.9184	0.9795	0.9966	0.9431	0.9191
dos	back	0.9973	0.7421	0.3212	0.7797	0.9510	1.0000
probe	satan	0.9945	0.9537	0.9846	0.9692	0.9138	0.5096
probe	nmap	0.9896	0.7655	0.9024	0.9001	0.4465	0.0547
r2l	warezclient	0.9941	0.1319	0.6843	0.5826	0.9490	0.0679

(3) модуль Neuronet производит разбор пакета для выделения информативных признаков;
(4) выделенные информативные признаки подаются на распознавание классификаторам;
(5) каждый из классификаторов производит анализ данных и возвращает в модуль Neuronet класс принадлежности пакета;
(6) модуль Neuronet возвращает в систему Snort значение, характеризующее пакет как аномальный/не аномальный;
(7) система Snort производит либо отсеивание пакета, либо передает его дальше по назначению.

Общая структура системы на базе IDS Snort и модуля Нейросетевого мониторинга сетевых атак Neuronet показана на рис. 3.

Структура модуля Neuronet выглядит следующим образом:

- `void SetupNeuronetPlugin(void)` {...} — регистрирует модуль как параметр в правилах системы Snort. Для обозначения любого модуля используется зарезервированное и идентифицирующее данный модуль ключевое слово. В случае данного модуля это слово «neuronet». В системе Snort данная функция вызывается на этапе связывания существующего модуля с ключевым словом. Это

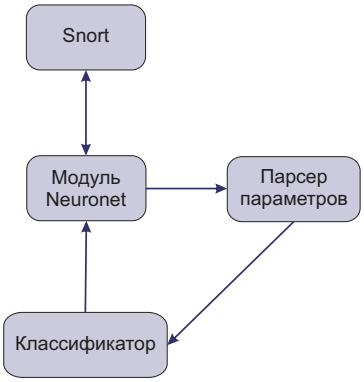

Рис. 3. Схема обучения и работы системы

достигается путем установления связи между ключевым словом и функцией инициализации модуля NeuronetInit;

- `static void NeuronetInit(char *data, OptTreeNode *otn, int protocol)` $\{\dots\}$ — инициализирует модуль. В качестве одного из параметров используются строка параметров к модулю, которая в ходе работы функции передается в качестве параметра функции `NeuronetRuleParse`. После этого происходит регистрация функции NeuronetDetect, которая и будет выполнять основную работу по распознаванию пришедших пакетов;
- `static void NeuronetRuleParse(char *data, OptTreeNode *otn, NEURONET_CTXT *nctx)` $\{\dots\}$ — позволяет разбирать описанные в правиле параметры модуля. Функция заполняет специальную структуру данных `NEURONET_CTXT`, являющуюся контекстом вызова функции проверки сетевого пакета NeuronetDetect;
- `int NeuronetDetect(void *context, Packet *p)` $\{\dots\}$ — позволяет проанализировать пришедший пакет, применяя имеющиеся настройки модуля. Производится анализ пришедшего пакета, и в случае выявления угрозы возвращается значение, сигнализирующее о необходимости отсеивания пакета;
- `int NeuronetCompare(void *l, void *r)` $\{\dots\}$ — производит сравнение параметров модуля;
- `uint32_t NeuronetHash(void *d)` $\{\dots\}$ — генерирует хэш-ключ исходя из параметров модуля.

9. Выводы

В настоящей работе предложена технология нейросетевого мониторинга сетевых атак с использованием IDS Snort. Разработанная на ее основе система показывает высокое качество распознавания ситуаций при работе с реальными потоками данных и может быть использована в составе различных программных комплексов для повышения уровня сетевой безопасности.

Работа выполнена в рамках Научно-технической программы Союзного государства «Разработка и использование программно–аппаратных средств ГРИД–технологий и перспективных высокопроизводительных (суперкомпьютерных)вычислительных систем семейства „СКИФ“».

Список литературы

[1] KDD Cup 1999 Data, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. ↑[]

[2] Саутин С. Н., Пунин А. Е., Савкович–Стевановић Е. Методы искусственного интеллекта в химии и химической технологии. Л. : ЛТИ, 1989. — 96 с. ↑1

[3] Buscema M. *Recirculation neural networks* // Substance Use and Misuse, 1998. Vol. **33**, no. 2, p. 383–388. ↑1

[4] Сапожников А. А. *Обнаружение аномальной сетевой активности* // Доклады Томского государственного университета систем управления и радиоэлектроники, 2009, № 1, с. 79–80. ↑2

[5] Русская группа пользователей Snort, http://www.snortgroup.ru/. ↑8

J. G. Emelyanova, A. A. Talalaev, I. P. Tishchenko, V. P. Fralenko. *Neural network technology of detection network attacks on information resources.*

Abstract. The article describes neural network approach, combining the speed of network traffic processing through compression characteristics and high efficiency of network attacks detection. Experimental results are presented.

Key Words and Phrases: artificial neural network, network attack, computing systems, feature mining, classification.

Образец ссылки на статью:

Ю. Г. Емельянова, А. А. Талалаев, И. П. Тищенко, В. П. Фраленко. *Нейросетевая технология обнаружения сетевых атак на информационные ресурсы* // Программные системы: теория и приложения : электрон. научн. журн. 2011. № 3(7), с. 3–15. URL: http://psta.psiras.ru/read/psta2011_3_3-15.pdf