УДК 81.93.29

К.С. Сумкин, А.Н. Тверской, Т.Ю. Морозова

Метод интеллектуальной поддержки принятия решений в задачах идентификации компьютерных атак

Предложен метод интеллектуальной поддержки принятия решений в задачах идентификации компьютерных атак, основанный на абдуктивном выводе, для нахождения причинно – следственных связей использована нейронная сеть Хопфилда. Получено решение задачи синтеза сложных гипотез, которое использует принцип параллельности нейронных сетей.

Ключевые слова: информационная безопасность, абдуктивный вывод, нейронные сети Хопфилда.

Защита телекоммуникационных систем от всех видов проникновения является новым направлением в области информационной безопасности. Эта задача стала особенно актуальной в последнее время в связи с возрастающим количеством сведений о разработке зарубежными странами концепций ведения информационной войны, т.к. они используются в сфере обороны, в экономике, транспорте, промышленности, связи, здравоохранении, при чрезвычайных ситуациях, в финансовых и банковских структурах, в области защиты и обеспечения правопорядка, а также спутниковой связи. Актуальность обеспечения информационной безопасности (ИБ) ТКС обусловлена высокими темпами роста (как качественного, так и количественного) компьютерных атак на объекты информатизации государственных структур РФ.

Каждый из уровней защиты успешно нейтрализует известные угрозы безопасности системы, однако оказывается малоэффективным при расширении поля угроз или обнаружении новых уязвимостей системы. ИБ ТКС все в большей степени обеспечивается за счет включения интеллектуальных средств в состав систем обнаружения компьютерных атак (СОА). Придание СОА таких качеств, как адаптивность и самоорганизация, свидетельствует о новом этапе развития средств автоматизации обеспечения ИБ ТКС. В работе предлагается задачу автоматизации обеспечения ИБ ТКС от компьютерных атак решать за счет включения интеллектуальных средств в состав средств защиты информации (СЗИ), а систему обнаружения компьютерных атак организовывать в виде адаптивной системы защиты. Объединение отдельных механизмов защиты в единый адаптивный комплекс, обладающий сведениями о состоянии защищаемой системы и происходящих в системе процессах, представляется актуальным [1].

Достоинством интеллектуальных средств защиты является наличие элементов самоорганизации и эволюции, которые используются для оперативных действий в СЗИ по классификации угроз и нейтрализации последствий вторжения. Общей чертой большинства существующих систем защиты информации является наличие средств идентификации атак (задача классификации) и оперативной реакции на несанкционированные проникновения в ТКС, а общим недостатком – отсутствие в системах защиты функций накопления и обобщения опыта взаимодействия ТКС с внешней средой и нейтрализацию угроз. Для успешного решения проблемы автоматизации обеспечения ИБ ТКС необходим комплексный подход и, прежде всего, иерархическая организация СЗИ с применением интеллектуальных средств для автоматической идентификации атак и накопления опыта нейтрализации угроз ИБ ТКС. При решении задач защиты информации интеллектуальные методы и средства позволяют учитывать профессиональный опыт экспертов ИБ, принимать решения в условиях неполной достоверности и искажения информации, адаптировать СЗИ к изменению угроз.

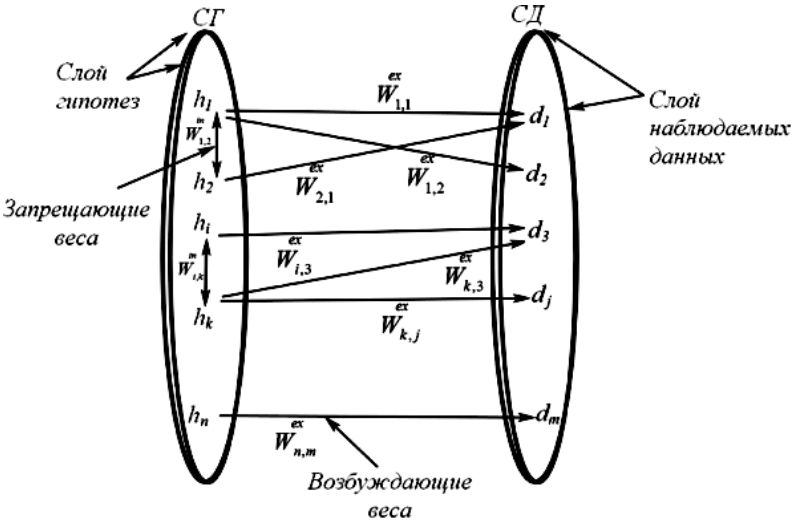
Основываясь на проведенном анализе существующих методов обучения систем, было предложено использовать абдуктивный вывод для решения задачи выявления причинно-следственных связей. Для реализации алгоритма выявления причинно-следственных связей при идентификации компьютерных атак использовались нейронные сети.

Анализ возможностей нейронных сетей показал, что они позволяют решать любые по сложности задачи абдукции и обеспечивают точные решения. Способность нейросетей к выявлению взаимосвязей между различными параметрами системы дает возможность выразить данные большой

размерности более компактно, если они тесно взаимосвязаны друг с другом. Обратный процесс представляет собой восстановление исходного набора данных из части информации.

С помощью проведенного анализа определений нейронных сетей и абдукции доказана возможность их совместного использования для решения задачи выявления причинно-следственных связей.

Абдукция может быть рассмотрена как обобщение из ряда наблюдений и синтеза гипотез для объяснения наблюдений. При имеющемся наборе гипотез предполагается, что алгоритм абдукции выберет одну из них, которая лучше всего объяснит наблюдаемые данные так, как она их понимает.



Для решения поставленной задачи представлена методика выявления причинно-следственных связей, использующая 2-слойную архитектуру (рис. 1). Здесь:

– 1-й слой состоит только из гипотез $h_1, h_2, ..., h_n$;

– 2-й слой состоит только из элементов наблюдаемых данных $d_1, d_2, ..., d_m$.

Рис. 1. Структура двухслойной сети

Эти два слоя связаны весами: $\overset{ex}{W}_{i,j}$ – возбуждающий вес; $\overset{in}{W}_{i,k}$ – запрещающий вес. Алгоритм работы данной модели основан на принципе конкуренции гипотез.

Для проверки работы алгоритма использован наглядный пример. Результаты отмечены на диаграмме, отображающей зависимость значений конкурирующих гипотез от различный временных интервалов (рис. 2).

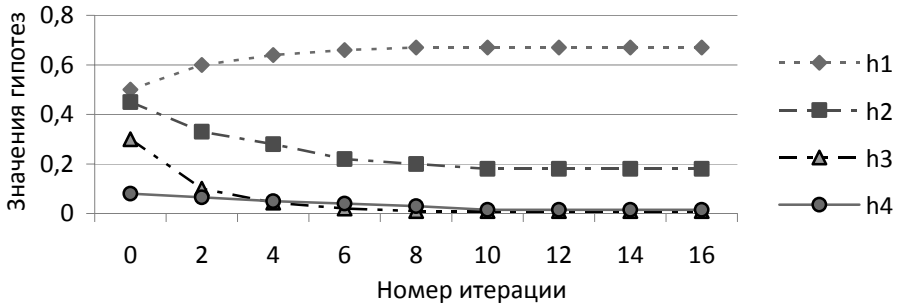


Рис. 2. Диаграмма зависимости значений гипотез от времени

Доказано, что одна гипотеза затухает и принимает значение, стремящееся к нулю, а три другие стабилизируются в значениях, отличных от нуля [2]. Принимая наибольшее значение среди рассматриваемых гипотез, h_1 является явным победителем и признается лучшим объяснением наблюдаемых данных. Гипотеза h_2 в условиях задачи определяется как неполная.

Результаты диаграммы говорят о том, что разработанный алгоритм позволяет выявить ложные или неполные гипотезы и повышает значение правильной, соответствующей гипотезы. Значения гипотез, получаемые в результате работы алгоритма, могут быть рассмотрены как их степени доверия.

Проверка работы данной модели на конкретных практических задачах показала, что при решении сложных задач абдукции, когда для объяснения всех элементов наблюдаемых данных недостаточно одной гипотезы, необходимо использовать третий слой СО – промежуточный.

Данный слой необходим для работы с объединением гипотез. Особенность этой нейронной архитектуры именно в наличии промежуточного слоя. Он используется лишь тогда, когда наблюдае-

мые данные можно объяснить только с помощью нескольких гипотез. Узлы на этом слое действуют как сложная гипотеза, представляющая собой объединение всех связанных гипотез.

Представленная трехслойная модель работает по тому же алгоритму, что и двухслойная, лишь с некоторыми изменениями в формулах.

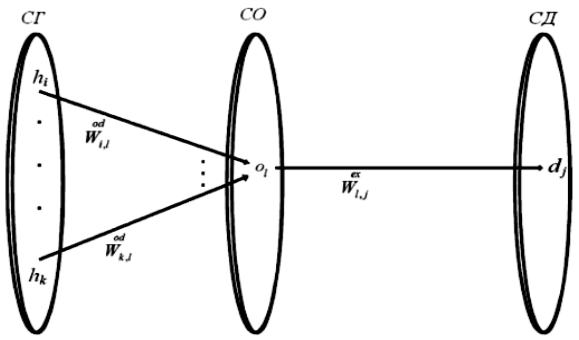


Рис. 3. Трехслойная нейронная архитектура

Метод выявления причинно-следственных связей использует нейронную сеть Хопфилда [3]. Выбор конкретной сети основан на результатах обзора свойств сети. Сеть Хопфилда позволяет просто и эффективно решать задачи воссоздания данных по неполной и искаженной информации. Для нахождения рационального решения задачи абдукции, определяется свойство минимальности составной гипотезы: составная гипотеза $\mathbf{H}_{c1}$ является лучшим объяснением наблюдаемых данных $\mathbf{D}_0$, нежели другая составная гипотеза $\mathbf{H}_{c2}$, если количество компонент первой меньше количества компонент второй, $|\mathbf{H}_{c1}| < |\mathbf{H}_{c2}|$. Формализована задача следующим образом: если $\mathbf{H}_c \in \mathbf{\Gamma}$, где $\mathbf{\Gamma} = \langle \mathbf{M}_{\text{оп}}, \mathbf{M}_{\text{пр}} \rangle$, то $\mathbf{H}_c \to \min$, т.е. учитывая, что $\mathbf{H}_c = col(h_1, h_2, ..., h_\Gamma)$,

$$\dim \mathbf{H}_c \to \min_{\mathbf{H}_c \in \mathbf{\Gamma}}.$$

Здесь $\mathbf{H}_c$ – подмножество множества гипотез $\mathbf{H}$, являющееся лучшим объяснением наблюдаемых данных $\mathbf{D}_0$, которое образует сложную (составную) гипотезу путем синтеза из набора простых гипотез $h_1, h_2, ...$; $\mathbf{\Gamma}$ – область допустимых простых гипотез; $\mathbf{M}_{\text{оп}}$ – максимальное объяснительное покрытие данных; $\mathbf{M}_{\text{пр}}$ – максимальное правдоподобие гипотезы; $\dim \mathbf{H}_c$ – вектор, элементы которого являются простыми гипотезами.

Доказано, что при решении задачи нахождения рационального решения абдукции могут возникнуть спорные ситуации между условиями области определения $\mathbf{\Gamma}$ и самой постановкой задачи. Для решения этой проблемы установлено отношение приоритета, в соответствии с которым максимальное покрытие данных имеет наивысший приоритет, а наличие минимального количества гипотез – наименьший [5].

Для решения задачи синтеза сложных гипотез было найдено средство реализации, которое обеспечило скорость работы в режиме реального времени. В качестве такого средства реализации предложено применять нейронную сеть, т.к. она использует принцип параллельности, что значительно влияет на скорость решения задачи.

Для решения поставленной задачи предложено преобразовать нейронную сеть Хопфилда в вычислительную модель нейронной сети. Для синтеза гипотез нейронные переменные $\mathbf{G}_j$ были связаны с каждой гипотезой $h \in \mathbf{H}_e$. Данная переменная дала возможность определять, включена ли простая гипотеза в составную.

Минимизировать количество элементов составной гипотезы $\sum_{j=1}^{m} \mathbf{G}_j$ удалось, введя следующее ограничение: все входные данные $d \in \mathbf{D}_0$ должны быть объяснены полностью, т.е. $\forall i = \overline{1,n}$,

$$\sum_{j=1}^{m} \mathbf{Q}_{ij} \mathbf{G}_j \geq 1,$$

где $\mathbf{Q}_{ij}$ – значения матрицы инцидентности, связывающей гипотезы и элементы данных; $\mathbf{G}_j$ – принадлежность простой j-й гипотезы к сложной: принимает значение 1, если j-я гипотеза принадлежит

сложной, и значение 0 в противном случае; $\mathbf{H} = \left\{ h_j \middle| \; j = \overline{1, M} \right\}$ – конечное множество элементарных (причинных) гипотез; $\mathbf{D} = \left\{ d_i \middle| \; i = \overline{1, N} \right\}$ – конечное множество входных данных (эффекты, факты, и т.д.); $\mathbf{H_e}$ – подмножество $\mathbf{H}$, в котором каждая гипотеза $h_j \in \mathbf{H}_e$ может объяснить некоторое непустое подмножество данных $\mathbf{D}_0$.

Учитывая, что элементы матрицы инцидентности $\mathbf{Q}_{ij}$ могут принимать значения 0 или 1, рассмотрим уравнение, определяющее степень покрытия данных:

$$P_{\mathbf{H}_{ck}} = \sum_{i=1}^{n} \prod_{j=1}^{m} \left\{ \left(1 - \mathbf{Q}_{ij}\right) + \left(1 - \mathbf{G}_j\right) \right\},$$

где $P_{\mathbf{H}_{ck}}$ – значение покрытия данных гипотезой $\mathbf{H}_{ck}$; k – количество возможных наборов гипотез, принимает значения от 1 до $\left(2^m - 1\right)$.

Это уравнение имеет следующие свойства:

1) каждое произведение не может быть отрицательным;

2) каждое произведение приравнивается к нулю, когда гипотеза, объясняющая исходную величину, входит в составную. Иначе, произведение принимает значение единицы;

3) уравнение принимает значение, равное нулю, когда определен состав гипотез, необходимых для объяснения всех элементов данных.

В качестве иллюстрации был рассмотрен пример, условия которого представлены на рис. 4.

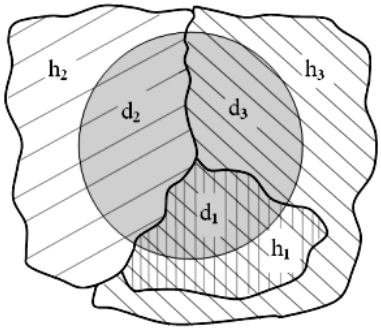


Рис. 4. Графическое представление объяснения трех элементов данных тремя гипотезами

По условию задачи элемент данных d_1 может быть объяснен как с помощью h_1, так и с помощью h_3, d_2 – только с помощью h_2, d_3 – только с помощью h_3. Необходимо выявить такую составную гипотезу, с помощью которой можно будет объяснить все три элемента данных, при этом решение должно быть рациональным. В данном примере продемонстрирован случай избыточного объяснительного покрытия, исключение которого очень важно на практике. Нахождение рационального решения позволит избежать дополнительных проверок для определения истинной атаки на ТКС [6].

Данный практический пример позволяет сделать вывод: для выделения рационального решения задачи абдукции необходимо, во-первых, исследовать возможные наборы элементарных гипотез на степень их объяснения наблюдаемых данных, а во-вторых, учесть количество этих гипотез, или компонент, в итоговом решении. Для этого введена величина S_k, которая представляет собой значение составной гипотезы.

Объединив оба условия для нахождения рационального решения задачи абдукции, формула вычисления значения составной гипотезы выглядит следующим образом:

$$S_k = \sum_{j=1}^{m} \mathbf{G}_j + P_{\mathbf{H}_{ck}} . \qquad (1)$$

Первое слагаемое в формуле определения значения составной гипотезы представляет собой количество элементов составной гипотезы, а второе слагаемое – значение покрытия данных, которое можно интерпретировать как величину штрафа за отсутствие полного покрытия, принимающую значение нуля в случае полного покрытия данных.

Вычислив значения составных гипотез для всех возможных наборов по формуле (1) [4], были получены результаты, представленные в таблице.

Основываясь на данных вычислениях, очевидно, что из всех возможных вариантов включения элементарных гипотез в сложную минимальное значение S_k соответствует гипотезе $\mathbf{H}_{ck} = \left(h_1, h_2\right)$, которая в данных условиях задачи является рациональным решением.

Значения составных гипотез

$\mathbf{H}_{ck}$	h_1	h_2	h_3	h_1, h_2	h_1, h_3	h_2, h_3	h_1, h_2, h_3
S_k	5	4	3	3	3	2	3

Экспериментальные исследования показали эффективность использования предложенной методики для решения задачи выявления причинно-следственных связей и идентификации компьютерных атак. Применение нечеткого извлечения знаний [7] позволит повысить эффективность использования нейронных сетей.

Литература

1. Зыков Д.Д. Проблема информационной безопасности производства наноэлектроники / Д.Д. Зыков, С.С. Бондарчук, Р.В. Мещеряков // Доклады ТУСУРа. – 2010. – № 1 (21), ч. 1. – С. 93-94.

2. Чень Ч. Математическая логика и автоматическое доказательство теорем / Ч. Чень, Р. Ли. – М.: Мир, 1983. – 360 с.

3. Kim C.S. A new fuzzy resolution principle based on the antonyms / C.S Kim., D.S. Kim, J.S Park // Fuzzy Sets & Systems. – 2000. – Vol. 113, issue 2. – P. 299–307.

4. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – СПб.: БХВ-Петербург, 2003. – 716 с.

5. Тверской А.Н. Разработка интеллектуальной системы поддержки принятия решений в задачах выявления причинно-следственных связей и нивелирования атак на компьютерную систему специального назначения // Спецтехника и связь. – 2011. – № 3. – С. 15–27.

6. Морозова Т.Ю. Система поддержки рассуждений в интеллектуальных системах / Т.Ю. Морозова, А.Н. Тверской // Юбилейный сборник научных трудов: под ред. А.П. Хныкина, А.Ю. Выжигина. – М.: МГУПИ, 2007. – С. 99–112.

7. Ходашинский И.А. Методы нечеткого извлечения знаний в задачах обнаружения вторжений / И.А. Ходашинский, И.В. Горбунов, Р.В. Мещеряков // Вопросы защиты информации. – 2012. – № 1. – С. 45–50.

Сумкин Константин Сергеевич
Канд. техн. наук, доцент каф. автоматизированных систем обработки информации и управления (АСОИУ) Московского государственного университета приборостроения и информатики (МГУПИ)
Тел.: 8 (926) 690-99-27
Эл. почта: skainet-1984@mail.ru

Тверской Антон Николаевич
Аспирант каф. АСОИУ МГУПИ
Тел.: 8 (926) 690-99-27
Эл. почта: tverskoy-anton@yandex.ru

Морозова Татьяна Юрьевна
Д-р техн. наук, проф. каф. АСОИУ МГУПИ
Тел.: 8 (916) 189-09-52
Эл. почта: tmorozova2006@rambler.ru

Sumkin K.S., Tverskoi A.N., Morozova T.Y.

Predictive decision support in identifying problems of computer attacks

Method for intelligent decision support tasks in the identification of the computer attacks based on Abductee Inference, to find the cause – effect relationships used by Hopfield neural network. Obtained by solution synthesis of complex hypotheses, which uses the principle of parallelism of neural networks.

Keywords: information security, abductee reasoning, Hopfield neural networks.