Г. В. Бабенко

ПРИМЕНЕНИЕ КАРТ САМООРГАНИЗАЦИИ КОХОНЕНА ДЛЯ ВЫЯВЛЕНИЯ ОТКЛОНЕНИЙ В ПОТОКЕ ТРАФИКА TCP/IP

Современный вектор развития общества тесно связан с процессом информатизации и совершенствования коммуникационных технологий. Телекоммуникационные системы внедрены практически во все сферы деятельности, исходя из этого и в результате постоянного расширения спектра разнообразных угроз информационной безопасности одной из основных особенностей корректного функционирования этих систем является обеспечение их необходимым уровнем защищенности. При этом контроль, обнаружение, реагирование на угрозы информационной безопасности в этих системах при сетевом взаимодействии являются одним из актуальных аспектов. Следовательно, для уменьшения вероятности успешной реализации угрозы необходимо иметь инструменты, позволяющие производить анализ информации о сетевых взаимодействиях и выявлять в ней признаки, способствующие успешной реализации угрозы.

В процессе обнаружения отклонений в функционировании сети обычно применяют либо сигнатурные — основанные на конечных характеристиках, либо поведенческие — основанные на штатном функционировании методы анализа [1]. Однако неспособность обнаруживать впервые реализованные злоумышленниками методы воздействия по характеристикам, заложенным в сигнатуре, и возникающие сложности в процессе корректного построения шаблона (правил) функционирования требуют использования метода с иной концепцией.

При реализации комплексного подхода, положенного в основу при разработке автоматизированной системы анализа сетевой инфраструктуры («АС2-И»), были применены в совокупности статистический, сигнатурный и нейросетевой методы анализа. Выбор данных методов обусловлен их «компонентонезависимостью» при анализе данных, что уменьшает вероятность наследования ошибок 1-го и 2-го рода [2]. Для решения задачи интеграции нейросетевого метода анализа применены самоорганизующиеся карты Кохонена (СОКК). Одним из основных аспектов, повлиявших на выбор именно этой технологии нейросетевого анализа, явилось то, что при формировании нейронной сети используется метод обучения «без учителя», т. е. результат зависит только от структуры входящей информации — свойств сетевого трафика. Эта характеристика СОКК существенно отличает данный метод от сигнатурного и статистического (поведенческого) методов анализа, так как анализ производится не с «шаблоном» (сигнатура или правила корректного функционирования), а с текущими характеристиками трафика. Отмеченный факт обеспечивает постоянство актуальной характеристики процессов, происходящих в сетевой инфраструктуре, а возможность преобразовать многомерные массивы характеристик трафика в

двумерные карты и отражать сравнительные отношения между характеристиками существенно облегчает процесс выявления подозрительных признаков. Одновременно, при сравнении с иными нейросетевыми алгоритмами, в аппарате СОКК отсутствие «учителя» сокращает время, затраченное на подготовку нейросети, так как формирование входных векторов характеристик и их подстройка под параметры самообучения происходят параллельно.

Таким образом, основной задачей, решаемой при использовании СОКК, является возможность нахождения закономерностей в сетевом трафике, свидетельствующих о потенциальных нарушениях информационной безопасности, и выявлении так называемых угроз «нулевого дня» [3] посредством построения и анализа карт признаков. Выбор протоколов семейства TCP/IP в качестве анализируемых обусловлен их всеобщим использованием при построении современных компьютерных сетей на основе Интернета. Необходимо также отметить, что в работе анализируются сетевые пакеты и метод адресации согласно стандарту IPv4.

Вопросы сбора, структуризации и алгоритм анализа данных

Применение в анализе СОКК основано на использовании сформировавшейся структуры нейронов, представляющих наборы векторов $w=[w_1, w_2, ..., w_n]^T$, где n определяется размерностью входных векторов. СОКК имеют прямую структуру распространения с одним вычислительным слоем, где все нейроны в решетке связаны со всеми узлами входящего слоя (рис. 1) [4].

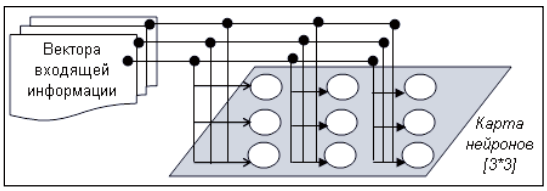


Рис. 1. Структура карты

Нейроны в ходе конкурентного процесса упорядочиваются по отношению друг к другу, создавая при этом на решетке систему координат. СОКК характеризуются формированием топографических карт входных образов, в которых пространственное местоположение нейронов решетки является индикатором встроенных статистических признаков, содержащихся во входящей информации [4]. Таким образом, решается основная задача применения СОКК – преобразование поступающей информации, имеющей произвольную размерность в двумерную дискретную карту [4].

Для анализа данных с использованием карт самоорганизации и формирования входных векторов из структурированной информации извлекаются характеристики, образующие следующие признаки (таблица 1).

Таблица 1. Пример входного вектора

Вектор характеристик пакета	**Источник**	**Получатель**	**Тип пакета**	**Порт источника**	**Порт назначения**	**Объем данных**	**Код ответа**
1	192.168.0.2	192.168.0.7	TCP/IP	8080	8080	16458	-
2	192.168.1.20	192.168.1.3	ICMP	-	-	32	9
…	…	…	…	…	…	…	…
N	x.x.x.x	y.y.y.y	TCP/IP	21	1432	16458	-

Принципы формирования входных векторов основаны на характеристиках информации, содержащейся в TCP/IP пакетах трафика. Ими являются: пары адресов источника и отправителя, пары портов назначения, по которым возможно более точно определить пакет, длину пакета, а также коды ответов для ICMP. При наличии вышеописанных характеристик существует возможность проанализировать историю сетевых взаимодействий и определить инциденты информационной безопасности.

Так как для проведения анализа требуется некоторое время, затрачиваемое на процесс накопления и структуризации информации, определим переменную ΔT — период проведения анализа. Исходя из состояния развития телекоммуникационных технологий ΔT рекомендуется устанавливать в диапазоне от 5 до 30 минут для детального анализа. Таким образом, имея значения ΔT — период проведения анализа, набор характеристик сетевого трафика — H, образующих входные вектора $x = [x_1, x_2, \ldots, x_m]^T$, а также функционал, реализуемый при построении карт самоорганизации — F, определим модель системы как кортеж $S = \sum : (\Delta T, H, F)$.

На начальном этапе алгоритма построения СОКК определяются синоптические веса нейронов. Для решения этой задачи применим генератор случайных чисел в интервале от 0 до 1. При входном векторе $x = [x_1, x_2, \ldots, x_m]^T$, вектор синоптических весов j-нейрона сети имеет вид $w_j = [w_{j1}, w_{j2}, \ldots, wj_n]^T$, при $j = 1,2,\ldots,l$, где l — количество нейронов в сети [5]. В данном случае имеется сеть Кохонена размером *[10*10]*, соответственно $l = 100$.

На этапе обучения карты важно подобрать вектор w_j, наиболее соответствующий входному вектору x. Для этого необходимо определить максимальные скалярные произведения — $\max(w_j^T x)$, что позволит выбрать то местоположение на карте, которое должно стать центром топологической окрестности возбужденного нейрона [5]. Так как наилучший критерий соответствия, основанный на максимизации скалярного произведения, — $\max(w_j^T x)$ — математически эквивалентен минимизации евклидова расстояния между векторами x и w_j, то, приняв нейрон-победитель *(best matching unit - BMU)* за w_c, получим $\|x - w_c\| = \min_j \{\|x - w_j\|\}$. Так как величина взаимодействия нейронов на карте определяется расстоянием между нейронами, то согласно данному фактору в работе выбор типа карты был сделан в пользу *гексагональной топологии*. Объяснимо это тем, что для шестиугольной сетки расстояние между нейронами больше совпадает с евклидовым, а это позволяет более точно определить нейрон-победитель [5].

После того как найден нейрон-победитель, необходимо откорректировать веса нейросети. Обозначим топологическую окрестность BMU — $h_{j,i}$, где i — нейрон-победитель, j — возбужденный нейрон. При этом латеральное расстояние между BMU и вторично возбужденным нейроном некое $d_{i,j}$ определяется как $d_{j,i}^2 = \left\| r_j - r_i \right\|^2$, где дискретный вектор r_j определяет позицию возбуждаемого нейрона, а r_i — позицию BMU. Таким образом, необходимо, чтобы $h_{j,i} = \max(h_{j,i})$ при $d_{i,j} = 0$ и $h_{j,i} = 0$ при $d_{i,j} \to \infty$. Для выполнения этих условий используем функцию Гаусса (функция 1):

$$h_{j,i(x)} = \exp\left(-\frac{d_{j,i}^2}{2\sigma(t)^2} \right), \tag{1}$$

где $\sigma(\tau)$ — эффективная ширина топологической окрестности. При этом $\sigma(t)$ является убывающей функцией от времени. Радиус обучения выбирается достаточно большим на начальном этапе обучения и постепенно уменьшается так, что в конечном итоге обучается один нейрон-победитель (функция 2):

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{T_1} \right), \tag{2}$$

где t — номер итерации,

σ_0 — начальное значение радиуса обучения,

T_1 — временная константа.

Так как в работе применяется двумерная решетка размером *[10*10]*, то для σ_0 возможно установить значение, равное радиусу решетки. Таким образом, $\sigma_0 = 10$. Соответственно, константа времени T_1 определяется по формуле 3:

$$T_1 = \frac{1000}{\log \sigma_0}. \quad (3)$$

Для успешного процесса самоорганизации карты — синоптической адаптации — необходимо, чтобы вектор синоптических весов w_j нейрона j изменялся в соответствии с входным вектором x. Для модификации весовых коэффициентов используется формула $w_j(t+1)=w_j(t)+\eta(t)h_{j,i(x)}(t)[x-w_j(t)]$, где $\eta(t)$ — параметр скорости обучения (функция 4):

$$\eta(t) = \eta_0 \exp\left(-\frac{t}{T_2}\right), \quad (4)$$

где t — номер итерации,

η_0 — начальное значение радиуса обучения,

T_2 — временная константа.

Параметр скорости обучения должен находиться в пределах от 0,1 до 0,01. Следовательно, в работе применим значения $\eta_0 = 0{,}1$ и $T_2 = 1000$. Рекомендуемое количество эпох обучения — более *500*[10*10] = 50000* итераций.

Визуализация и расшифровка результатов применения карт

Для решения задачи визуализации результатов СОКК применим линейный градиент стандарта RGB. С привязкой к среде разработки — Visual Studio 2010, язык C# 4.0, получим следующие цветовые сочетания и их комбинации (таблица 2):

Таблица 2. Значения «цветовой» переменной

<table>
<tr><th>Цвет</th><th>Фиксированные значения</th><th colspan="2">Комбинации переменных</th></tr>
<tr><td>Красный — max</td><td>255.0.0</td><td rowspan="3">Ближе к max</td><td>green = 255 – ((mas[i, j] – ((average + max) / 2)) * 255) / (max – ((max + average) / 2))</td></tr>
<tr><td>Оранжевый</td><td>255.255.0</td><td rowspan="2">red = ((mas[i, j] – average) * 255) / (((max + average) / 2) – average)</td></tr>
<tr><td>Желтый</td><td>0.255.0</td></tr>
<tr><td colspan="4">mas[i, j] — характеристика нейрона в сетке размером [i*j], «green» — значение, характеризующее зеленый цвет, «average» — среднее значение, «red» — красный цвет</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td>Зеленый</td><td>0.255.0</td><td>Среднее</td><td></td></tr>
<tr><td colspan="4"></td></tr>
<tr><td>Синий — min</td><td>0.0.255</td><td rowspan="3">Ближе к min</td><td>green = ((mas[i, j] – min) * 255) / (((average + min) / 2) – min)</td></tr>
<tr><td>Фиолетовый</td><td>255.0.255</td><td rowspan="2">blue = 255 – ((mas[i, j] – ((average + min) / 2)) * 255) / (average – ((average + min) / 2))</td></tr>
<tr><td>Красный — max</td><td>255.0.0</td></tr>
<tr><td colspan="4">mas[i, j] — характеристика нейрона в сетке размером [i*j], «green» — значение, характеризующее зеленый цвет, «average» — среднее значение, «blue» — синий цвет</td></tr>
</table>

Цвет	Фиксированные значения	Комбинации переменных
Белый	*255.255.255*	
Черный	0.0.0	

При использовании вышеописанного метода каждая карта является характеристикой признаков входящих векторов, при этом цвета ячейки определяются на основе значения компоненты mas[i, j]. Те ячейки, в которые попали элементы с минимальными значениями компоненты, отображаются темным цветом, с последующим осветлением при возрастании значений компоненты (рис. 2).

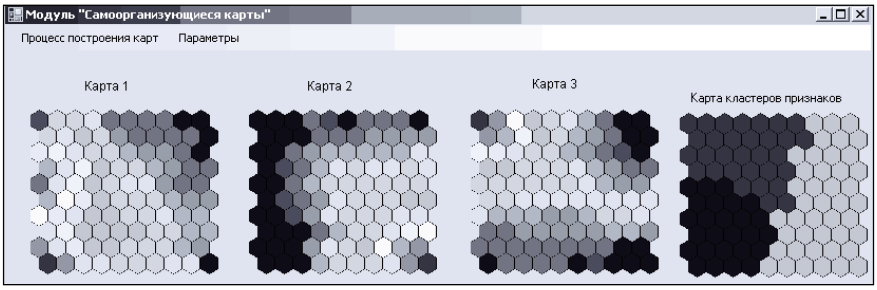


Рис. 2. Карты трех признаков и блоки кластеров

Вектора, расстояние между которыми внутри группы меньше, чем расстояние до соседних групп, образуют кластер признака группы. Вычислив расстояние между вектором весов нейрона и его ближайшими соседями, определим кластера признаков.

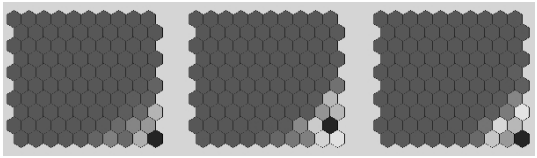
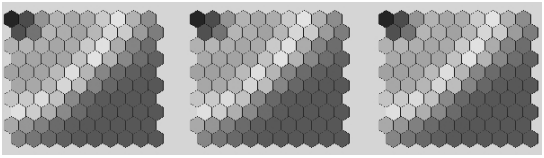
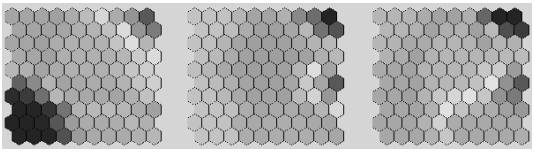
Сценарии (наборы) карт напрямую зависят от состояния сети, так как СОКК «подстраивается» под закономерности, имеющиеся во входных данных, и пригодны для анализа безопасности сети на основе объективных характеристик. При активации режима анализа «АС2-И» автоматически переводит сетевой адаптер в «promiscuous mode», выступая в качестве снифера пакетов. Аппарат СОКК, извлекая характеристики захваченных снифером пакетов, выделяет на картах подозрительные признаки в сетевой инфраструктуре, воспользовавшись чем администратор может определить периоды аномальной активности участников сетевого взаимодействия.

При проведении процедур анализа рекомендуемыми характеристиками входных векторов являются: «адреса», «тип пакета», «порты (протоколы)». При таком наборе данных существует возможность оценить сетевую инфраструктуру на наличие инцидентов информационной безопасности, согласно технологии передачи пакетов. Так, идентифицировав источник подозрительной активности — признак «адрес», необходимо последовательно проанализировать связанные характеристики: ими чаще всего выступает «тип пакета» (TCP/UDP с IP-заголовком или ICMP-пакет), далее — идентифицировать службу активации — «протокол (порт взаимодействия)». При наличии контролируемой политики распределения сетевых адресов (DHCP) возможно выявить источник, определяемый СОКК и воздействовать на него, что позволит существенно снизить вероятность успешной реализации угрозы безопасности. Алгоритм анализа может иметь иные шаги, в зависимости от визуальной оценки карты.

Имея на картах признаков общую характеристику всех анализируемых объектов, можно визуально определить отличающиеся блоки по отношению к соседним, что практически не требует временных затрат. При проведении периодического или постоянного анализа состояния сетевой инфраструктуры появляется возможность идентифицировать действия злоумышленников, направленные на подготовку к проведению атак: сканирование портов, службы, получение

информации об используемом прикладном, системном программном обеспечении и системах защиты (на этапе рекогносцировки). Четко зафиксировать время начала негативного воздействия на объект атаки (вторжение) из-за его краткосрочности достаточно затруднительно, однако определить наличие последующего воздействия на объект при помощи аппарата СОКК является возможным, после чего по отношению к источнику угрозы необходимо применять иные меры воздействия. В итоге на построенных картах признаков может быть отображена информация об изменениях в долях используемых протоколов (служебных, прикладных, пользовательских), фиксируются изменения в нагрузке на узлах сети (ICMP-Flood, запросы), по количеству зафиксированных пакетов определяется объем переданной/полученной информации, регистрируется теневое использование ресурсов сети и т. п. (таблица 2).

Таблица 2. Примеры чтения карт признаков

Отображение карты	**Расшифровка**
	Детектирование пакетов ARP (3), посылаемых DNS (2) сервером сети, при отсутствии иной сетевой нагрузки (ЭВМ входит в состав домена).
	Характеристика признаков сетевого адаптера в период с 18.00 до 8.00 — равномерное распределение (ЭВМ не входит в состав домена).
	Детектирование чрезмерной активности по обмену TCP-пакетами (1), источника (2), по порту 80 — протокол HTTP (3).

Стоит отметить, что корректная расшифровка карт вручную требует базовых знаний в области стека протоколов TCP/IP. В «АС2-И» реализован механизм автоматической расшифровки карт, в котором одновременно анализируются 5 из 7 описанных характеристик, однако наиболее информативным является визуальный анализ администратором системы. Также необходимо отметить, что использование аппарата СОКК не позволяет предотвратить инициирование атаки, основной задачей его применения является идентификация предпосылок угрозы безопасности, а также информирование о происходящих в период анализа подозрительных процессах.

Заключение

Таким образом, после определения ключевых характеристик потока сетевого трафика в работе были сформированы входные вектора, которые использовались при реализации алгоритма построения СОКК. С использованием линейного градиента был разработан способ визуализации полученных карт. Расшифровка полученной совокупности карт признаков позволяет проводить анализ взаимодействий в сетевой инфраструктуре, а реализованный в «АС2-И» автоматический режим чтения карт позволяет анализировать сети на заданном временном интервале. Аппарат СОКК возможно применить для поиска скрытых закономерностей в трафике, идентификации аномалий сетевого трафика и детектирования деятельности «инсайдеров» в совокупности с иными, широко распространенными системами защиты информации (антивирусные системы, межсетевые экраны, системы обнаружения вторжений и атак, системы анализа защищенности) для повышения уровня защищенности.

СПИСОК ЛИТЕРАТУРЫ:

1. *Леонтьев В. П.* Безопасность в сети Internet. М.: ОЛМА Медиа Групп, 2008. — 256 с.
2. *Бабенко Г. В.* Систематизация методов анализа сетевых взаимодействий // Наука: поиск-2010: сб. научн. ст. Астрахань: изд-во АГТУ, 2010. С. 56—58.
3. *Радько Н. М., Скобелев И. О.* Риск-модели информационно-телекоммуникационных систем при реализации угроз удаленного и непосредственного доступа. М.: РадиоСофт, 2010. — 232 с.
4. *Кохонен Т.* Самоорганизующиеся карты. М.: Бином. Лаборатория знаний, 2008. — 655 с.
5. Самоорганизующиеся карты Кохонена — математический аппарат. [2008]. URL: http://www.basegroup.ru/library/analysis/clusterization/som (Дата обращения: 23.01.2011).