АНАЛИЗ ВХОДНЫХ ДАННЫХ ДЛЯ НЕЙРОСЕТЕВОЙ СИСТЕМЫ ОБНАРУЖЕНИЯ АТАК В РАЗЛИЧНЫХ СЕТЕВЫХ ОКРУЖЕНИЯХ

Кочурко П. А., БГТУ, Брест

Системы обнаружения атак (СОА) используются для обнаружения различных типов атак. Они объединяются с межсетевыми экранами и другими средствами обеспечения безопасности для того, чтобы своевременно оповещать персонал в случае обнаружения подозрительной активности. На текущий момент в обнаружении атак используются различные технологии [1], в том числе и искусственные нейронные сети (ИНС), которые могут применяться на разных этапах обнаружения атак: в качестве детектора аномалий [2], в качестве детектора злоупотреблений или распознавания типа атаки [3], на этапе предобработки данных для уменьшения размерности входных данных и др.; для различных целей применяются и различные архитектуры ИНС.

В качестве входных данных для анализа сетевой активности наиболее часто берутся журналы регистрации [1] или непосредственно данные сетевого трафика [1-3]. В любом случае, вопрос выделения признаков для анализа решается в итоге чаще всего одинаково: и записи журналов регистрации, и данные трафика конвертируются в записи о соединениях, анализ которых значительно эффективнее, чем анализ пакетов. Параметры, которые могут подаваться на вход СОА, варьируются в зависимости от целей и технологии работы детектора и делятся на следующие группы:

- внутренние параметры соединения – такие как длительность работы, количество переданных байт, порты, флаг результата и т. д.;
- параметры данных – количество попыток и отказов регистрации в системе, shell-запросов и т. п.;
- статистические параметры – количество соединений с данным сервисом, количество запросов от данного хоста в течение последних n секунд и т. п.

ИНС в качестве детекторов атак [2-3] применялись следующим образом: на небольшой выборке из базы данных KDD нормальных соединений и/или соединений-атак обучались соответствующие ИНС, после чего, благодаря способности к обобщению и функционированию в окружении с большим количеством шумов, они становятся способны обнаруживать атаки во всех соединениях в данной базе, причём процент ошибок зачастую значительно меньше, чем при применении других методов обнаружения атак. Однако при анализе реального сетевого трафика таким обученным детектором процент ложных срабатываний неожиданно многократно возрастает.

Для выяснения причины данного явления проанализируем параметры, подающиеся на вход сетей в [3]: длительность работы соединения, количество переданных байт в обе стороны, флаг результата соединения, тип протокола, сервис, флаг регистрации в системе (logged in). Шесть из семи параметров относятся к внутренним параметрам соединений, последний – параметр данных. Могут ли различаться данные параметры для сходных по природе нормальных соединений в различных сетевых окружениях настолько, чтобы быть принятыми за атаку?

Рассмотрим два одинаковых TCP соединения, которые установлены с одним и тем же удаленным сокетом с хостов в разных сетевых окружениях. Из перечисленных выше

у них точно не должны различаться типы протокола, сервис, флаг регистрации. При одинаково успешном (или не успешном) результате работы соединений флаг результата тоже должен быть одинаковым. Разниться же могут количества переданных байт (но не сильно), и, главным образом, длительность работы соединения, например, вследствие различий в скорости и технологии подключения к сети Интернет или в скорости передачи данных внутри сети. Статистический анализ данных входных параметров и результатов показывает, что данные параметры наиболее существенно влияют на качество обнаружения атак.

Как можно избежать проблем, связанных с различными значениями параметров сходных по природе соединений в различных сетевых окружениях? Поставлен следующий эксперимент: на трёх хостах в различных сетях с разной скоростью доступа в Интернет (dial-up, Ethernet, dial-up) запрашивались из веб-браузера подряд шесть URL (протокол – tcp, сервис – http): http://ibrest.net; http://dynamo.brest.by; http://santa-bremor.com; http://mail.tut.by; http://santa-bremor.com/products/surimi/,ru,; http://www.bstu.by; http://iit.bstu.by. Естественно, что результат, который был получен в браузере, ничем друг от друга не отличался. Сниффер (разработанный на технологии WinPCap) же выдал следующие параметры (таблица 1).

Как видно из таблицы, идентичные действия в различных сетевых окружениях приводят к различной сетевой активности, вплоть до различных количеств соединений и пакетов. Сравним длительности работы соединений в разных сетях (Рис. 1).

Таблица 1 – Результаты работы

	Пакетов	TCP	UDP	ICMP	Соед.
Dial-up 1 (2)	2866	2787	78	1	177
Dial-up 1 (3)	2728	2650	77	1	179
Ethernet (4)	2757	2757	0	0	144
Ethernet (5)	2721	2721	0	0	149
Ethernet (6)	2783	2779	4	0	156
Dial-up 2 (7)	2249	2197	52	0	87
Dial-up 2 (8)	2278	2234	28	0	92

Как видно из рисунка 1, длительности работы нормальных соединений одинакового характера в разных сетевых окружениях серьёзно варьируются. Кроме того, даже одинаковые соединения в одной сети работают с разной продолжительностью, хотя и более сходно, чем в разных сетях.

Для того, чтобы получить сходные значения параметров соединений одинаковой природы перед подачей на вход СОА их необходимо статистически нормировать:

$$x_i^1 = (x_i^0 - M_i) / \sigma_i . \quad (1)$$

Таким образом, учитывая среднее значение длительности работы соединения, мы в какой-то мере учитываем среднюю скорость передачи данных в данной сети и от рассмотрения параметра «длительность работы соединения» переходим к параметру «нормированное отклонение от средней длительности». Рисунок 2 показывает, что значения в различных средах становятся намного более сходными.

Аналогичный анализ для количества переданных и полученных байт показывает, что нормирование данных параметров не влияет на степень подобия в различных средах – ненормированные и нормированные значения дают примерно одинаковую картину.

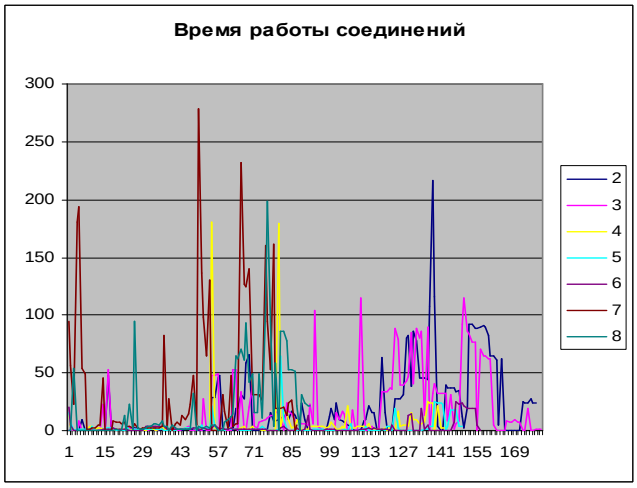

Рисунок 1 – Длительности работы соединений

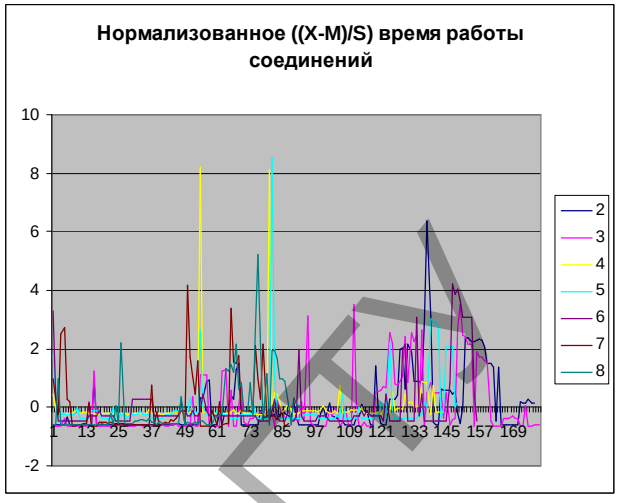

Рисунок 2 – Нормированные длительности

Литература

1. S. T. Brugger. Data Mining Methods for Network Intrusion Detection. http://www.bruggerink.com/ ~zow/Projects.html

2. П. Кочурко. Нейросетевой детектор аномалий. Известия Белорусской инженерной академии, № 1(19)/2'2005 – с. 78-81.

3. Vladimir Golovko, Pavel Kochurko. *Intrusion Recognition Using Neural Networks.* In Proc. of IDAACS'2005, September, Sofia, Bulgaria, 2005

НЕЛИНЕЙНАЯ МНОГОСЛОЙНАЯ НЕЙРОННАЯ СЕТЬ В ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ ПОТРЕБЛЕНИЯ ЭЛЕКТРОЭНЕРГИИ.

Кочурко Ю.В., БГТУ, Брест

Введение

Предсказание потребления электроэнергии является актуальной задачей и играет ключевую роль в технико-экономическом функционировании объектов энергосистемы. Так, владение предварительными данными о нагрузке, с экономической стороны, позволяет значительно усовершенствовать тарифную политику для объектов энергопотребления и, тем самым, снизить коммерческие потери, а с технической – обеспечивает экономный и безопасный режим работы энергосистемы.

Начиная с 1990 года, активно рассматриваются возможности применения нейронных сетей для решения задачи предсказания нагрузки путем прогнозирования. В настоящее время имеется множество научных публикации, в которых рассматривается возможность прогнозирования нагрузок с помощью нейрокомпьютеров [1-3]. Также рассматривается задача построения краткосрочных предсказаний нагрузок с повышенной точностью. Исследована релевантность нескольких известных моделей. Предложен новый метод прогнозирования, основанный на использовании трехслойных искусственных нейронных сетей с комбинированной структурой, объединяющих линейные и нелинейные схемы.

1. Описание нейронной сети для решения задачи прогнозирования

Для решения задачи прогнозирования потребления электроэнергии использовалась нелинейная многослойная нейронная сеть (многослойный персептрон), поскольку ее