Инженерный вестник Дона, №4 (2018)
ivdon.ru/ru/magazine/archive/n4y2018/5355

Применение интеллектуального анализа данных для обеспечения безопасности веб-сайтов

А. Мансур, Ж. Мохаммад, В.В. Галушка

Донской государственный технический университет, Ростов-на-Дону

Аннотация: *Статья посвящена актуальной проблеме обеспечения информационной безопасности web-сайтов. В ней рассматривается способ обнаружения преднамеренных угроз конфиденциальности информации, возникающих в результате несанкционированного доступа и проявляющихся в виде обращений к ресурсам, нетипичных для конкретного пользователя. В статье предлагается метод, основанный на интеллектуальном анализе данных. Его суть заключается в классификации поведения пользователей на основе информации о совершаемых им действиях с использованием искусственной нейронной сети. Описывается её структура, метод обучения и способы практического применения, даётся оценка эффективности предложенных методов.*

Ключевые слова: интеллектуальный анализ данных, искусственные нейронные сети, web-дизайн, машинное обучение, классификация.

Введение

Важным механизмом защиты веб-сайтов от угроз безопасности является обнаружение аномальной или нетипичной активности, которая может являться следствием действий злоумышленников, направленных на нарушение целостности, конфиденциальности или доступности информационных ресурсов, в совокупности составляющих основу понятия информационной безопасности [1, 2]. Рассматриваемый далее метод защиты основан на применении интеллектуального анализа данных с использованием искусственных нейронных сетей для выявления на сайтах нетипичной активности с целью повышения точности её выявления при условии большого числа обращений к ресурсам сайта.

1. Модель представления исходных данных

Большинство компьютерных систем записывают события, связанные с данной системой, в файлы журналов [3]. В правильно

настроенном веб-сервере любые действия, являющиеся нарушением безопасности, также приводят к появлению соответствующей записи в одном или нескольких файлах журнала. Аналогичным образом могут функционировать средства аудита безопасности сайта, добавляя в определённую таблицу базы данных, необходимые сведения об обращениях к ресурсам сайта [5]. Указанная таблица должна содержать следующий минимальный набор полей:

— userID — идентификатор пользователя, который является внешним ключом таблицы “users”;
— object — объект, с которым пользователь выполнял действие (таблица БД, либо файл или папка);
— action — действие, например, чтение или запись, если рассматривать файл, или добавление, удаление, обновление, если рассматривать таблицу БД;
— timestamp — время совершения действия.

2. Метод определения нетипичных обращений к ресурсам веб-сайтов

Перечисленные поля и их конкретное содержание в таблице журнала будут служить исходными данными для дальнейшего анализа действий пользователей и выявления потенциальных злоумышленников путём классификации шаблонов поведения на «нормальные» и «подозрительные».

Критерием определения подозрительного поведения является количество обращений пользователя к одному или всем имеющимся ресурсам в течение определенного периода времени, или частота его использования [6]. В общем случае, когда количество действий, которые пользователь выполняет на ресурсе, превышает определенный предел,

отличный от обычного, поведение этого пользователя можно считать подозрительным.

Большое количество исходных данных, а также сложность их внутренней структуры затрудняют применение для анализа известных статистических методов, в связи с чем перспективным является метод, основанный на искусственных нейронных сетях [7].

Их применение в данном случае сводится к задаче классификации [8] активности каждого пользователя, решение которой заключается в обучении нейронной сети распознавать шаблоны поведения на основе данных об обращениях пользователя к ресурсам системы. Полученные шаблоны сохраняются и используются при анализе новых действий. Если действия пользователя не соответствуют сохраненным естественным шаблонам, поведение этого пользователя будет считаться подозрительным.

3. Применение искусственных нейронных сетей для решения задачи классификации пользовательской активности

Основываясь на описанных ранее принципах программной реализации искусственных нейронных сетей и используя разработанный набор классов [9], можно создать нейронную сеть для классификации пользовательской активности. Она будет состоять из трех слоёв — входного, скрытого и выходного, количество нейронов, в каждом из которых необходимо определить.

Входные векторы искусственной нейронной сети представляют собой строки таблицы, полученные в результате процесса предобработки исходных данных описанных выше. Каждый вектор выражает поведение пользователя за определенный период времени и представлен в виде промежуточных статистических значений, полученных путём применения агрегатных функций языка SQL при выборке исходных данных из таблицы журнала. Так,

например, при 5 ресурсах, с каждым из которых можно выполнить одно из 3 действий: получение, обновление и создание нового экземпляра, количество элементов входного вектора будет составлять 15. Количество нейронов в выходном слое определяется предметной областью, то есть исходя из необходимости различать два класса, количество нейронов выходного слоя будет равно 1, а его выход будет интерпретироваться в соответствии со следующим правилом: [0, 0.5] — нормальный, (0.5, 1] — подозрительный. Число нейронов скрытого слоя определяется эмпирически, исходя из опыта и предварительных практических результатов, и в данном случае равняется десяти.

После формирования нейронной сети её можно обучить выполнению процесса классификации [10]. При этом для обучения «нормальному» поведению используются данные, имеющиеся в журнале и выбранные за период, в котором не было нарушений безопасности сайта, а для обучения «подозрительному» — генерируются на основе критериев, сформулированных выше. Обучение нейронной сети происходит с использованием алгоритма обратного распространения ошибки и останавливается, когда ошибка достигает наименьшего допустимого значения.

Заключение

Нейронная сеть, реализованная и обученная в соответствии с описанным методом, позволяет проводить классификацию поведения пользователей сайта на основе анализа данных о совершаемых ими действиях по отношению к имеющимся ресурсам, полученных из таблицы журнала, заполняемого подсистемой аудита безопасности. Её применение позволит оперативно выявлять потенциальных злоумышленников, а также повысит точность классификации поведения пользователей за счёт интеллектуального анализа данных.

Литература

1. Роднин А.В., Турчик В.Ю. Концепция применения интеллектуального анализа данных в средствах защиты информации баз данных // Физика. Технологии. Инновации: сборник научных трудов. Екатеринбург: УрФУ, 2015. С. 263-269.

2. Liu B. Web Data Mining Exploring Hyperlinks, Contents, and Usage Data. Second Edition. Springer. 2011. 622 p.

3. Min L. Application of Data Mining Techniques in Intrusion Detection. An Yang Institute of Technology. 2005. pp.1273-1277.

4. Daniel B., Sushil J. Applications of data mining in computer security. Springer Science & Business Media. 2002. 252 p.

5. Lambert I., Glenn M. Security Analytics: Using Deep Learning to Detect Cyber Attacks. University of Norths Florida. 2017. 83 p.

6. Аникин И.В. Технология интеллектуального анализа данных для выявления внутренних нарушителей в компьютерных системах // Научно-технические ведомости СПбГПУ. 2010. №6 (113). С. 112–117.

7. Новиков Ф.А. Дискретная математика для программистов. С.-Пб.: Питер, 2000. 364 с.

8. Мансур А., Мохаммад Ж., Галушка, В.В. Применение методов интеллектуального анализа данных при разработке системы классификации компетентностей студентов для web-сайта университета // Инженерный вестник Дона, 2018, №2. URL: ivdon.ru/ru/magazine/archive/N2y2018/4838.

9. Али М., Галушка В.В. Особенности применения средств объектно-ориентированного программирования для реализации многослойных искусственных нейронных сетей прямого распространения // Автоматизация: проблемы, идеи, решения. Уфа: Омега Сайнс, 2017. С. 107-109.

10. Галушка В.В., Фатхи В.А. Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз

данных // Инженерный вестник Дона, 2013, №2. URL: ivdon.ru/ru/magazine/archive/n2y2013/1597/.

References

1. Rodnin A.V., Turchik V.YU. Fizika. Tekhnologii. Innovacii: sbornik nauchnyh trudov. Ekaterinburg: UrFU, 2015. pp. 263-269.

2. Liu B. Web Data Mining Exploring Hyperlinks, Contents, and Usage Data. Second Edition. Springer. 2011. 622 p.

3. Min L. An Yang Institute of Technology. 2005. pp.1273-1277.

4. Daniel B., Sushil J. Springer Science & Business Media. 2002. 252 p.

5. Lambert I., Glenn M. Security Analytics: Using Deep Learning to Detect Cyber Attacks. University of Norths Florida. 2017. 83 p.

6. Anikin I.V. Nauchno-tekhnicheskie vedomosti SPbGPU. 2010. №6 (113). S. 112–117.

7. Novikov F.A. Diskretnaya matematika dlya programmistov. [Discrete Mathematics for Programmers]. S.-Pb.: Piter, 2000. 364 p.

8. Mansur A., Mohammad ZH., Galushka, V.V. Inženernyj vestnik Dona (Rus), 2018, №2. URL: ivdon.ru/ru/magazine/archive/N2y2018/4838.

9. Ali M., Galushka V.V. Osobennosti primeneniya sredstv ob"ektno-orientirovannogo programmirovaniya dlya realizacii mnogosloynyh iskusstvennyh neyronnyh setey pryamogo rasprostraneniya. Avtomatizaciya: problemy, idei, resheniya. Ufa: Omega Cayns, 2017. pp. 107-109.

10. Galushka V.V., Fathi V.A. Inženernyj vestnik Dona (Rus), 2013, №2. URL: ivdon.ru/ru/magazine/archive/n2y2013/1597/.