УДК 004.056.53

БАЙЕСОВСКИЕ СЕТИ КАК СРЕДСТВО ПРЕДСТАВЛЕНИЯ СЦЕНАРИЕВ ПОВЕДЕНИЯ ВРЕДОНОСНЫХ ПРОГРАММ

А. В. Молдавская, к. т. н. В. М. Рувинская

Одесский национальный политехнический университет
Украина, г. Одесса
amme4od@mail.ru

Исследование посвящено проблеме автоматизированного формирования сценариев, описывающих поведение вредоносных программ. Предложена модель для их представления в виде байесовских сетей и метод машинного обучения для ее формирования. Применимость метода для формирования сценариев опробована на примере.

Ключевые слова: вредоносные программы, сценарии, байесовские сети

Для обнаружения вредоносных программ существуют два основных способа анализа: статические и динамические. Статические методы изучают текст программы, а динамические – программы во время их работы, а наибольшая их эффективность достигается при совместном использовании [1]. В настоящее время растет интерес к динамическим методам анализа, в том числе, поведенческим. Поведение многих новых вредоносных программ остается типичным: например, это пересылка данных пользователя, генерация трафика, саморассылка и самокопирование [2]. Рост объемов вредоносного программного обеспечения, появляющегося ежедневно, ставит задачу автоматизации процесса поведенческого анализа. Существующие средства (эмуляция в виртуальных средах, генерация отчетов по использованию API-функций и системных вызовов и т. п.) обеспечивают лишь частичную автоматизацию труда эксперта [3]. Ранее в [4] было предложено обнаруживать вредоносные программы экспертной системой, построенной на основе предварительно сформированных сценариев поведения.

Целью работы является выбор методов машинного обучения, которые подходили бы для формирования сценариев поведения вредоносных программ на основе примеров. Как следствие, стоят задачи выбора модели, наиболее подходящей для представления сценариев, выбора метода машинного обучения для нее и, наконец, опробование этого метода при построении сценариев поведения вредоносных программ.

Сценарий представляет собой модель представления знаний, описывающую поведение как обобщенную последовательность действий [4]. Для описания поведения одной вредоносной программы достаточно получить список ее действий во время выполнения и задать связи между ними. Нашей задачей является получение сценариев, построенных на основе наблюдений за множеством таких программ, которые будут вмещать в себе различные варианты их поведения для разных типов. Выбрана наиболее подходящая модель — байесовская сеть, поскольку она является направленным графом (выражает причинно-следственные связи) и содержит вероятности (позволяет осуществлять вероятностный вывод).

Процесс построения байесовской сети с помощью некоторого алгоритма называется структурным обучением. Для опробования предложенной модели и способа ее обучения выбран алгоритм структурного обучения байесовских сетей K2, так как он наиболее подходит для обработки большого количества данных. Для проверки применен обобщенный сценарий поведения почтовых «червей», описанный Лабораторией Касперского [5] (рис.1). Для обучения была использована выборка по поведению почтовых «червей» из энциклопедии SecureList от Лаборатории Касперского [6]. Также эта выборка была дополнена данными о том, как ведут себя «черви» в случае, когда им не удалось осуществить одно из своих действий (часть прочих действий также не может быть после этого осуществлена). Это позволило значительно полнее отразить в выборке связь некоторых действий друг с другом, то есть повысить ее репрезентативность.

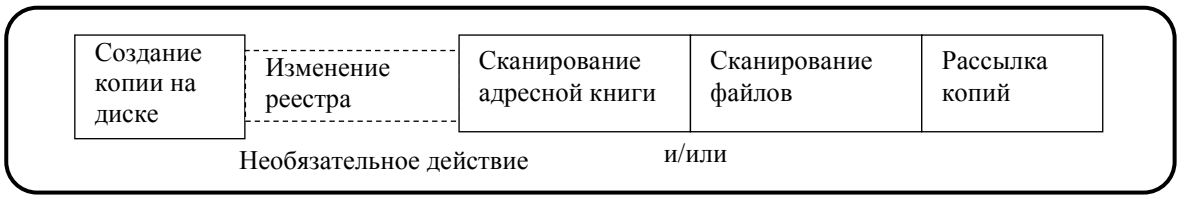

Рис. 1. Сценарий типичного поведения почтового «червя»

Результатом обучения стал нижеследующий граф (рис. 2).

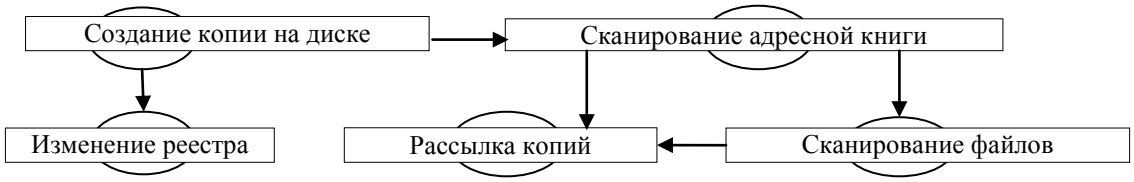

Рис. 2. Результат структурного обучения сети

Поясним этот результат. Тот факт, что из изменения реестра не следует ни один из следующих узлов, говорит о том, что изменение реестра является необязательным событием в сценарии, что соответствует экспертной версии. Сканирование файлов действительно следует из сканирования адресной книги: если «червь» не сканировал адресную книгу, вероятность того, что он будет сканировать файлы, повышается, так как из файлов тоже можно извлекать адреса электронной почты. Рассылка копий следует как из сканирования адресной книги, так и из сканирования файлов по той же причине: адреса можно получить из обоих источников. Сходящаяся связь аналогична отношению «и/или». Таким образом, сеть построена верно и полностью соответствует сценарию.

Для представления сценария была выбрана байесовская сеть. Опробована пригодность алгоритма структурного обучения K2 для формирования сети, соответствующей сценарию поведения почтового «червя». Результат соответствует проверочному образцу. Дальнейшее направление исследований – разработка методики поведенческого анализа с помощью сценариев, методики машинного обучения сценариев.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. A.Moser, C. Kruegel, E. Kirda. Limits of Static Analysis for Malware Detection // Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual.– P. 421— 430

2. Honig A. Practical Malware Analysis.– No Starch Press, 2012.– 800 p.

3. M. Egele, T. Scholte, E. Kirde, Ch. Kruedel. A Survey on Automated Dynamic Malware Analysis Techniques and Tools [Электронный ресурс] / Режим доступа: https://www.seclab.tuwien.ac.at/papers/malware_survey.pdf

4. Рувинская В. М. Эвристические методы детектирования вредоносных программ на основе сценариев / В. М. Рувинская, Е. Л. Беркович, А. А. Лотоцкий // Искусств. интеллект .– 2008.– № 3.– С. 197—207.

5. Вирусы и средства борьбы с ними. Учебный курс. [Электронный ресурс] // Режим доступа: http://www.csid.omsu.omskreg.ru/docs/docs/viruses.pdf

6. SecureList — Новые описания детектируемых объектов. [Электронный ресурс] // Режим доступа: http://www.securelist.com/ru/descriptions

---

A. V. Moldavskaya, V. M. Ruvinskaya

Usage of Bayesian networks to represent malware behavior scenarios.

This paper describes the problem of automated building of decision scenarios which can represent the behavior of malicious software. We propose to represent scenarios using Bayesian networks and build them using machine learning methods. The experiment results are presented.

Keywords: *malware, scenarios, machine learning, Bayesian networks, structure learning.*