из различных источников. Таким образом, появляется возможность оперативного выявления угроз безопасности, регистрации инцидента и проведения соответствующих действий по обнаружению нарушителя и предотвращению атак.

Нарушения политики безопасности могут иметь серьезные последствия для организации, такие как потеря репутации на рынке, огромные финансовые затраты. Поэтому организации нуждаются в быстром выявлении угроз безопасности и реагировании на возникшие нештатные ситуации. TSOM способен помочь предотвратить вторжения, повысить уровень безопасности в организации и существенно облегчить труд администраторов ИБ.

СПИСОК ЛИТЕРАТУРЫ:

1. http://www-01.ibm.com/software/tivoli/products/security-operations-mgr/.
2. Документация IBM Tivoli Security Operations Manager.

А. Д. Чорняк

СТАТИСТИЧЕСКИЙ МЕТОД ВЫДЕЛЕНИЯ СИГНАТУР РАЗРУШАЮЩИХ ПРОГРАММНЫХ ВОЗДЕЙСТВИЙ

Одним из основных методов борьбы с вирусами является сигнатурный анализ [1]. Однако отсутствие жестко заданного алгоритма выделения и методики оценки сигнатур, а также наличие случайных ошибок, возникающих при ручном анализе, снижают надежность данного подхода. Таким образом, возникает необходимость создания автоматизированной системы выделения сигнатур разрушающих программных воздействий (РПВ). В данной работе предлагается статистический подход к автоматизации процесса выделения сигнатур.

Для начала необходимо выделить множество файлов, подверженных РПВ (R). С этой целью производится запуск модели РПВ, после чего осуществляется сравнение контрольных сумм файловой системы до и после заражения. Элементы множества R можно получить не только исполнением самого РПВ, но и запуском его потомков. Чем больше множество R, тем меньше вероятность ошибок типа «false negative», а это является наиболее важным фактором при сигнатурном анализе [1].

Далее необходимо создать максимально полное множество всех программ, не подверженных РПВ. Это множество будет представлять собой «чистую систему» (С). Чем оно больше, тем меньше вероятность ошибок типа «false positive».

По сути, задача выявления сигнатур сводится к нахождению последовательности, присутствующей во всех файлах, зараженных данной моделью РПВ, и не встречающейся ни в одном файле «чистой системы». Т. е. необходимо найти новое множество потенциальных сигнатур (SP), представляющее собой вычитание множества С из множества R.

Любую из полученных последовательностей можно использовать в качестве сигнатуры. Однако то, что выбранные последовательности не встречаются в «чистой системе», еще не гарантирует отсутствие ложных срабатываний. Это связано с тем, что невозможно составить полное

множество «чистой системы» C, так как оно будет стремиться к бесконечности. Данная задача разрешима только для узких областей применения, где набор возможного программного обеспечения ограничен [2]. Следовательно, необходимо провести дополнительный этап анализа множества SP с целью получения последовательности, представляющей собой качественную сигнатуру. Здесь под качественной сигнатурой понимается последовательность, обеспечивающая минимальное количество ложных срабатываний.

Качественная сигнатура должна содержать как можно больше эвристических признаков, характерных для моделей РПВ. Исходя из этого предлагается создать множество эвристических признаков (E), каждому из которых поставить в соответствие некоторый «вес» (ρi, $i = 1,k$, где k — число эвристических признаков), который будет соответствовать вероятности ложного срабатывания в случае наличия данного признака в сигнатуре.

При обнаружении нескольких эвристических признаков в оцениваемой сигнатуре предлагается применить метод, основанный на байесовском подходе к вычислению вероятностей. Суть его состоит в следующем. Проверке подлежит некоторая гипотеза H, с которой связаны утверждения $A1$, $A2$ и т. д. Тогда, проверив одно из них (A), мы можем пересчитать вероятность гипотезы H с учетом истинности или ложности этого утверждения.

В рамках данной работы в качестве гипотезы H будет выступать предположение о том, что файл, содержащий оцениваемую сигнатуру, является инфицированным либо самим РПВ. В качестве утверждений (A) предлагается использовать эвристические признаки, присутствующие в сигнатуре.

Таким образом, будут получены конечные вероятности выдвинутой гипотезы для каждой выделенной сигнатуры: $\rho 1(H/A1\ A2...As)$, $\rho 2(H/A1\ A2...As)$ … $\rho i(H/A1\ A2...As)$, где s — количество эвристических признаков в сигнатуре. Из вычисленных вероятностей берется максимальная, так как она обеспечит наименьшую вероятность ложного срабатывания.

Оценка полученных вероятностей в соответствии с заданным уровнем значимости α осуществляется на основе анализа вхождения параметра ρi в доверительный интервал.

$$\overline{P} - t\frac{\sigma}{\sqrt{n}} < P < \overline{P} + t\frac{\sigma}{\sqrt{n}}, \qquad (1)$$

где $t\frac{\sigma}{\sqrt{n}} = \delta$ — точность оценки, σ — среднеквадратическое отклонение, n — объем выборки, $\overline{\rho}$ — выборочное среднее, t — аргумент функции Лапласа, при котором $\Phi(t) = \frac{\alpha}{2}$.

Среднеквадратическое отклонение рассчитывается как:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - \overline{P})^2}. \qquad (2)$$

В данном случае выборка будет включать в себя совокупность всех весов ρi, $i = 1,k$, а также все вариации их произведений.

Вместо множества эвристических признаков в описанном выше методе можно также использовать частотные признаки. Для этого рассчитывается распределение значений байт по частоте их использования в базе данных известных сигнатур (или вирусов), которое принимается как эталонное. «Веса» ставятся в соответствие каждому значению байт, отражая частоту появления последнего в базе сигнатур. Чем больше частота, тем больше «вес». Далее производится поиск значений байт из эталонного распределения во множестве потенциальных сигнатур SP. Если в элементе множества SP содержатся несколько байт, значения которых совпадают со значениями из эталонного распределения, то «веса» суммируются. Следовательно, из множества SP выбирается сигнатура с минимальным значением ρ.

Если вероятность ложного срабатывания у всех полученных сигнатур достаточно велика, то для исключения ошибки типа «false positive» можно совместно использовать пару потенциальных

сигнатур. При этом общая вероятность ложного срабатывания будет равна произведению вероятностей каждой сигнатуры. Однако недостатком данного подхода является увеличение времени сканирования вдвое.

Эффективность применения описанного метода обусловлена следующими факторами:

- использование жестко заданных алгоритмов позволит существенно повысить скорость выявления сигнатур;
- исключение человеческого фактора сведет к минимуму случайные ошибки;
- автоматизация процесса позволит использовать методы, требующие больших временных затрат;
- использование аппарата теории вероятности позволит дать оценку полученным сигнатурам.

СПИСОК ЛИТЕРАТУРЫ:

1. *Касперски К., Рокко Е.* Искусство дизассемблирования. СПб.: БХВ-Петербург, 2008. — 891 с.
2. *Cohen F.* Computer viruses // Computers & Security. 1987. № 6. P. 22–35.