V. S. Matveeva

The Criterion for Assessing the File Content for Its Proximity to the Random Data

Keywords: random data, statistical tests, encrypted data, density of distribution, assessment of uniformity, localization of heterogeneity, wavelet transform

The article is focused on the search of the random data in the file system, which is an important step in digital forensics in case of detecting encrypted data. Encryption is widely used today, as well as by malefactors to conceal data, suggesting that encrypted information can be hidden in the file system. There are means of detection of encrypted files, but they have significant limitations. Statistical tests can also be used for this purpose but they are resource and time consuming. Therefore, the article proposes a new criterion to searching for random data, based on the assessment of uniformity of data on the special plane, with the ability to localize the heterogeneity, which reduces the type I error to zero.

В. С. Матвеева

КРИТЕРИЙ ОЦЕНКИ СОДЕРЖИМОГО ФАЙЛОВ РАЗЛИЧНЫХ ФОРМАТОВ НА ПРЕДМЕТ ИХ БЛИЗОСТИ К СЛУЧАЙНЫМ ДАННЫМ

Поиск файлов со случайными данными может осуществляться при проведении компьютерных исследований и судебных компьютерно-технических экспертиз с целью выявления скрываемых зашифрованных данных, зашифрованных вредоносных компонентов программ. В работе под зашифрованными данными подразумеваются данные, статистические свойства которых близки к свойствам случайных чисел [1], которые, в свою очередь, понимаются как выходная последовательность генераторов псевдослучайных чисел.

Такие файлы в первую очередь обладают высокой энтропией. Так как расчет энтропии основывается на вероятностях возникновения символов, то есть не учитывается внутренняя структура файла, значения, близкие к 8, могут принимать не только файлы со случайными и зашифрованными данными, но и файлы сложных форматов данных: видео- и аудиофайлы, файлы-архивы, исполняемые файлы и т. д. [2, 3, 4].

Кроме этого, для выявления случайных данных могут использоваться статистические тесты, применение которых целесообразно только в совокупности для перекрытия ошибок друг друга. При этом задействуется значительное количество ресурсов для поиска файлов с зашифрованными данными и времени на подсчет статистических значений в каждом тесте.

Все существующие тесты применяются на основании представления файла f в виде конечной последовательности байт $\{x_1, x_2, \ldots, x_N\}$ длины N, где $x_i \in X = \{0, 1, \ldots, 255\}$, $|X| = k = 256$ — мощность множества X, а N — размер файла. Такая последовательность анализируется байт за байтом или по блокам слева направо для проверки равномерности и независимости элементов последовательности относительно друг друга.

Рассмотрим эту последовательность иначе. Нанесем последовательность на плоскость, по оси абсцисс которой отложены значения от 0 до N, а по оси ординат — от 0 до 255. Заполним эту плоскость точками P(x,y), такими, что x — номер байта в последовательности байт файла, а y — значение этого байта.

Логично предположить, что для случайных данных точки на этой плоскости будут нанесены равномерно. Это следует из того, что подпоследовательность случайной последовательности также случайна, а значит, имеет равномерное и независимое распределение по всем своим значениям. Так как элементы наносятся на плоскость в соответствии с их положением в последовательности

и значением, то на этой плоскости они должны располагаться равномерно. Пример фрагмента такой плоскости приведен на рис. 1.

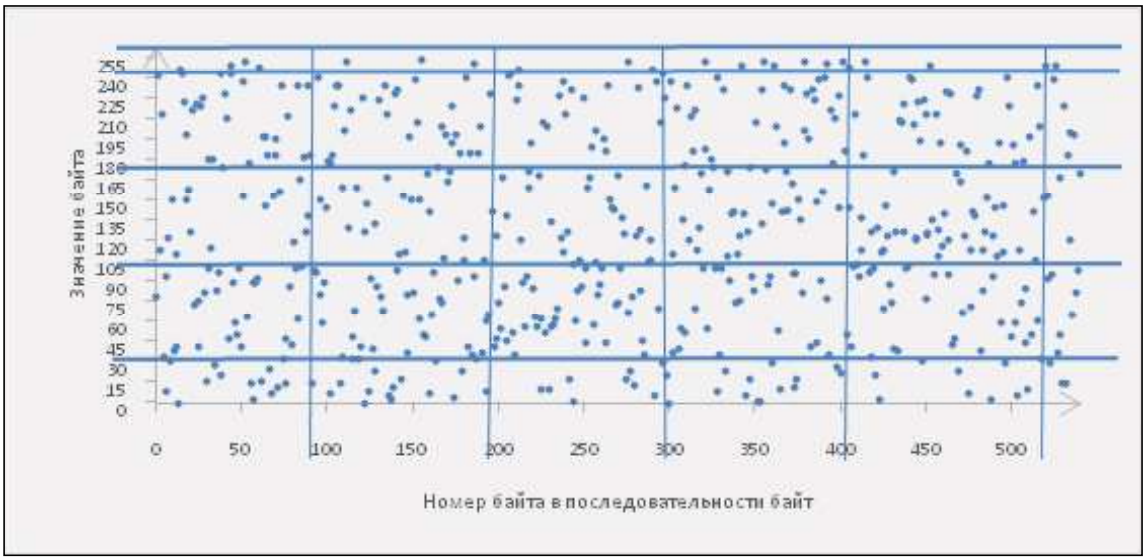

Рис. 1

Разобьем эту плоскость на фрагменты размером W x H. W и H задаются произвольно, с помощью метода скользящего окна.

Если распределение на плоскости равномерное, то и плотность распределения элементов в каждом окне одинакова.

Для каждого фрагмента производим подсчет плотности распределения точек в фрагменте относительно размера фрагмента, то есть

$$\rho_i = \frac{(\text{количество точек из файла в } i-\text{м фрагменте})}{(\text{количество всех точек в } i-\text{м фрагменте} = W * H)},$$

где $i \in [1,L]$, L — количество получившихся фрагментов в результате прохождения по содержимому плоскости методом скользящего окна размером W на H.

Осуществляя подсчет плотности каждого фрагмента, можно сформировать вектор плотностей, то есть выполнить преобразование:

$S: f \rightarrow \{\rho_1, \rho_2, \ldots, \rho_L\}$.

В рамках тестирования форматов файлов с высокой энтропией, не говоря уже об остальных форматах файлов, выявлено, что в распределении плотностей имеются выраженные всплески, которые и предлагается отслеживать для выявления зашифрованных данных.

Поэтому в работе предлагается подход для выявления таких всплесков на основании вейвлет-анализа. Для этого выдвигается и проверяется **гипотеза**:

Значения модулей вейвлет-коэффициентов, рассчитанных для случайных или зашифрованных данных, не превышает значения ε, которое получается экспериментально.

Проверка гипотезы заключается в проверке **критерия:**

Если вейвлет-преобразование последовательности плотностей для файла удовлетворяет требованию:

$$|W_{a\,b}| - \varepsilon > 0, \forall W_{a\,b},$$

где W_{ab} — вейвлет-коэффициент с параметром сдвига b и параметром масштаба a, а ε определяется экспериментально, то файл признается не содержащим данные, близкие к случайным.

Тестирование подхода показывает нулевую ошибку первого рода и возможность локализации отклонений в распределении байт в файле.

СПИСОК ЛИТЕРАТУРЫ:

1. *Giordano J., Maciag C.* Cyber Forensics: A Military Operations Perspective // International Journal of Digital Evidence. 2002. V. 1. I. 2. P. 1–13.
2. *Jozwiak I., Kedziora M., Melinska A.* Theoretical and Practical Aspects of Encrypted Containers Detection – Digital Forensics Approach // Dependable Computer Systems. 2011. P. 75–85.
3. *Weston P., Wolthusen S.* Forensic Entropy Analysis of Microsoft Windows Storage Volume // Information Security for South Africa. 2013. P. 1–7.
4. Digital Forensics File Carving Advances [Электронный ресурс]: KoreLogic DFRWS-2006 Project. 2006. URL: http://www.korelogic.com/Resources/Projects/dfrws_challenge_2006/DFRWS_2006_File_Carving_Challenge.pdf.

REFERENCES:

1. *Giordano J., Maciag C.* Cyber Forensics: A Military Operations Perspective // International Journal of Digital Evidence. 2002. V. 1. I. 2. P. 1–13.
2. *Jozwiak I., Kedziora M., Melinska A.* Theoretical and Practical Aspects of Encrypted Containers Detection – Digital Forensics Approach // Dependable Computer Systems. 2011. P. 75–85.
3. *Weston P., Wolthusen S.* Forensic Entropy Analysis of Microsoft Windows Storage Volume // Information Security for South Africa. 2013. P. 1–7.
4. Digital Forensics File Carving Advances: KoreLogic DFRWS-2006 Project. 2006. URL: http://www.korelogic.com/Resources/Projects/dfrws_challenge_2006/DFRWS_2006_File_Carving_Challenge.pdf.

A. P. Mikhalkova, A. S. Zaytsev

The Baysean Approach Appliance for Early Detection of Insider Threat

Key words: insider, Bayesian network

Purpose of report is to increase the information security level of an organization from insider threats. The report presents the method of detection and determination of the type of supposed insider threat with the help of the Bayesian approach.

А. П. Михалькова, А. С. Зайцев

О ПРИМЕНЕНИИ БАЙЕСОВСКОГО ПОДХОДА ДЛЯ РАННЕГО ОБНАРУЖЕНИЯ ВНУТРЕННИХ НАРУШИТЕЛЕЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

В настоящее время существует множество исследований, посвященных противодействию внешним угрозам ИБ, разработаны полноценные методологии и специализированное программное обеспечение. Вопрос противодействия внутренним угрозам ИБ проработан не настолько хорошо, но в то же время 77,6 % руководителей ИТ и ИБ заявляют, что основная опасность для организаций связана с внутренними угрозами, а именно с утечкой информации ограниченного доступа, нелояльным или преступным поведением сотрудников и пр. [1]. Основа успешного противодействия инсайдеру – его раннее выявление. Ввиду этого крайне актуальной задачей является разработка эффективного метода обнаружения внутренних нарушителей на раннем этапе реализации угрозы ИБ.