УДК 004.492.3

ИСПОЛЬЗОВАНИЕ ХЕШ-ФУНКЦИЙ ДЛЯ КЛАССИФИКАЦИИ И ОБНАРУЖЕНИЯ ВРЕДОНОСНЫХ ПРОГРАММ

Рыбаков С.Д. (Университет ИТМО, г. Санкт-Петербург)

Аннотация. В работе рассматривается решение задачи распознавания и определения взаимосвязей между различными экземплярами вредоносных программ путем использования принципов скользящего хеширования и вычисления хеша импортируемых библиотек.

Введение. В связи с развитием компьютерной техники, люди используют все большее количество устройств, обладающих самым разным функционалом. Закономерно, что это приводит к уверенному увеличению числа злоумышленников, желающих воспользоваться несовершенством систем безопасности или человеческим фактором.

Одним из наиболее часто встречающихся видов преступления является использование вредоносных программ. Однако они написаны людьми и поэтому зачастую обладают всеми присущими обычным компонентам ПО свойствами: переиспользованием кода, заимствованиями из других проектов и зависимостью от конкретных библиотек. Это дает возможность обнаружения вредоносного ПО, путем сравнения с уже известными образцами, построения зависимостей или классификации семейств малвари и примерного определения «эволюционного пути» экземпляра, что может быть полезным при расследовании атаки.

Основная часть. Для решения задачи анализа данных удобно воспользоваться криптографическими хеш-функциями. Наиболее популярные, например, SHA-1, MD5 выгодно применять для поиска точных соответствий, но при любом изменении входных значений, они вычисляют новые, существенно отличающиеся от предыдущих, результаты.

Поэтому существуют также и «нечеткие», кусочные хеши. Наиболее популярным среди них является Context triggered piecewise hashing (CTPH, fuzzy hashing), то есть контекстно вызываемое кусочное хеширование. Этот принцип использован в алгоритме SSDeep, разработанном для компьютерной криминалистики. Он позволяет вычислять традиционные хеши для частей файла различного размера, для чего используется механизм скользящего окна. Это дает возможность находить в файлах схожие или одинаковые блоки при помощи расчёта расстояния Дамерау-Левенштейна, то есть минимального количества операций редактирования данных, необходимого для получения из текущей последовательности заданной. Таким образом, становится возможным выявлять измененные вредоносные файлы, если они были основаны на ранее уже использованных и обнаруженных экземплярах.

Менее эффективным «в вакууме» является хеш импортируемых библиотек, поскольку он производит хеширование только используемых библиотек и связанных с ними функций, а их теоретически можно скрыть, хотя бы частично. Однако на практике этим пользуются также далеко не все, и, поэтому можно выявить вредоносное ПО, стремящееся получить функционал тех же библиотек, что и уже выявленные ранее экзмепляры.

Использование этих двух способов хеширования совместно позволяет получать существенно лучшие результаты анализа данных, так как они компенсируют взаимные недостатки и находят наиболее значимые следы, оставляемые человеческим фактором, в коде вредоносных программ.

Выводы. В работе на практике подтверждена эффективность совместного использования результатов вычисления кусочных хешей и хешей импортированных библиотек для определения взаимосвязей или зависимостей между различными экземплярами вредоносного программного обеспечения.

Рыбаков С. Д. (автор) Подпись