А. Е. Антонов, ассистент филиала ФГБОУ ВПО «Национальный исследовательский университет «МЭИ» в г. Смоленске

А. С. Федулов, докт. техн. наук, профессор, директор филиала ФГБОУ ВПО «Национальный исследовательский университет «МЭИ» в г. Смоленске

Идентификация типа файла на основе структурного анализа

Необходимость идентификации объектов возникает при решении многих прикладных задач, в частности, задачи выявления вредоносного ПО. Получить нужный результат можно сравнивая структуру файла с известными подозрительными программами.

Введение

В ряде задач, возникающих при использовании современных информационных технологий, требуется сравнение файлов или их частей. К таким задачам относятся, например, классификация, обнаружение незаконного использования данных, поиск дублирующихся участков программного кода. К одной из важных практических задач, для решения которой может проводиться сравнение файлов, относится автоматическое обнаружение вредоносного программного обеспечения (ПО). Современной компьютерной вирусологии известно множество вредоносных программ, каждая из которых представляет собой незначительную модификацию одного и того же исполняемого кода. Это значительно усложняет работу антивирусных программ, основанных на сигнатурном анализе: фактически необходимо проанализировать каждую из модификаций.

Классификация вредоносных программ на семейства, каждое из которых содержит похожие файлы, представляющие собой модификации одного и того же кода, значительно облегчает борьбу с вирусами, так как позволяет для установления факта принадлежности программы какому-либо семейству проводить анализ, используя толь-

ко один или небольшое число представителей этого семейства. Такая идентификация может быть выполнена на основе автоматического сравнения файлов и обнаружения среди них похожих.

В данной работе предложен способ сравнения файлов на основе анализа их структуры.

Обзор существующих подходов к сравнению файлов

Сравнение исполняемых вредоносных файлов рассматривается в работах ряда авторов. Анализ этих работ показал, что в качестве исходных данных для сравнения могут выступать следующие сущности исполняемого файла: набор и последовательность функций интерфейса операционной системы [1–3], строковые константы [4, 5], статистические свойства [5, 6], структура [7, 8]. Результатом сравнения является некоторая мера близости, подобия (*similarity measure*) файлов. Значения меры подобия лежат в промежутке [0, 1]. Малые значения меры соответствуют схожим файлам, большие — различным.

Под структурой файла в настоящей работе понимается взаимное расположение частей файла, например, секций, ресурсов, а также расположение блоков данных внутри частей.