**УДК 004.02**

# ИДЕНТИФИКАЦИЯ ИСПОЛНЯЕМЫХ ФАЙЛОВ ПРИ ИХ ПЕРЕДАЧЕ ПО СЕТИ

**Салахутдинова К.И.** (Университет ИТМО)

Исследование направлено на рассмотрение хорошо известной проблемы мониторинга несанкционированной установки программного обеспечения из сети пользователями. Обсуждается подход к идентификации исполняемого файла, с применением анализа его побайтового представления и формирования неполных частотных сигнатур в соответствии с принципами передачи файлов по сети и MTU.

**Введение.** В ряде случаев возникает необходимость решения задач идентификации, верификации и валидации свободно распространяемого программного обеспечения. Многие решения в основном ориентированы на отслеживание фиксированного состояния программного кода на носителях и в оперативной памяти, что не всегда позволяет быстро определять разрешенные модификации и изменения версий. Такие ученые, как Казарин О. В., Копыльцов А. В., Сорокин И. В., Корнблюм Дж. Д., Лонг С., Эбрингер Т., Сун Л., Бозтас С., Шульц М. Г., Эскин Е., Сантос И. и другие посвятили свои работы различным подходам к идентификации программного обеспечения. Однако существующие подходы к идентификации установленного программного обеспечения имеют ряд ограничений и недостаточных возможностей для обеспечения комплексной реализации мер информационной безопасности.

**Основная часть.** Автором было проведено множество исследований в этой области и логичным продолжением работы является разработка динамической идентификации исполняемых файлов при их передачи по сети. Данная задача предполагает идентификацию файла на основе неполного владения им, а точнее, рассмотрение ситуации получения (загрузки) файла на компьютер пользователя и идентификацию его на основе полученного числа фрагментов. Пусть у нас есть некоторый набор исполняемых файлов $F$. Также пусть каждый из этих файлов принадлежит некоторой программе $N_i$, тогда ее версии будут составлять набор $N_i = \{f_1,f_2,...,f_m\}$. Тогда у нас есть $i$ различных программ, каждая из которых может быть представлена $m$ различными версиями. Представим каждый исполняемый файл $f_m$ в виде упорядоченной последовательности характеристик, которую назовем сигнатурой файла $S_{fm} = <s_1,s_2,...,s_k>$. Сигнатура $S_{fm}$ представляет собой вектор в $k$-мерном пространстве. Пусть у нас есть сигнатуры не известных нам программ $S_{fm}$, которые мы хотим идентифицировать и сигнатуры известных программ $S_N$, которые мы будем использовать при сравнении. Задачей в данном случае будет являться построение такого алгоритма идентификации, который способен определить меру сходства идентифицируемого исполняемого файла $f_m$ с именем $N_i$ определенной программы, которая удовлетворяла бы следующему критерию: точность (метрика для оценки результатов идентификации) должна быть максимальной.

**Выводы.** Автором был проведен эксперимент по формированию сигнатур исполняемых файлов при их передачи по сети и исходя из результатов, можно однозначно сказать, что идентификация не вредоносного программного обеспечения, передаваемого по сети, имеет право на существование. Были получены хорошие результаты, позволяющие распознать исполняемый файл с точностью 72,5%, имея только 95% его кода. Причем точность достигает 80,3% при наличии 99% его кода. В то же время стоит отметить объем изучаемых файлов. Минимальный размер исполняемого файла составлял 5,5 КБ, а максимальный-7,5 МБ.