«МЕТОД РАСПОЗНАНИЯ ВРЕДОНОСНЫХ ФЛЭШ-САЙТОВ, ТРЕБУЮЩИХ ОТПРАВИТЬ SMS»

А.Д. Скороходов
Московский государственный институт электроники и математики (технический университет)

АННОТАЦИЯ

В статье дается обзор и применение способов распознания вредоносных флэш-сайтов, требующих от посетителей отправления SMS на короткий номер. Рассматривается собственное решение данной проблемы. Приводится описание технологии Flash.

ВВЕДЕНИЕ

Основная цель работы – разработка алгоритма распознавания вредоносных флэш-сайтов с просьбой отправить СМС и исследование возможных путей по улучшению производительности полученного алгоритма.

В ходе работы был разработан алгоритм распознавания вредоносных флэш-сайтов с просьбами отправить SMS на короткий номер. Данный алгоритм позволяет предостеречь пользователей интернета от подобных сайтов, предоставив им уведомление о том, что данный сайт несёт в себе угрозу безопасности. Алгоритм также предусматривает добавление URL таких сайтов в базу данных Dr.WEB. В разработанном алгоритме используется алгоритм распознавания текстов в файлах SWF.

Разрабатываемый алгоритм позволяет решить такие проблемы защиты информации в сети Интернет, как:

- распознание флэш-сайтов со скрытой угрозой и предупреждение пользователей о возможной опасности;
- пополнение базы данных Dr.WEB ссылками на вредоносные флэш-сайты содержащие просьбу об отправлении SMS;
- борьба со злоумышленниками в сети Интернет, использующих сервисы оплаты посредством SMS.

В статье описываются технологии распознания текстов в SWF файлах, приводится результат поиска аналогичных решений, приводятся качественные характеристики разработанного алгоритма и пути дальнейшего развития проекта.

ТЕХНОЛОГИЯ FLASH

Adobe Flash позволяет работать с векторной, растровой и ограниченно с трёхмерной графикой, а также поддерживает двунаправленную потоковую трансляцию аудио и видео. Для КПК и других мобильных устройств выпущена специальная "облегчённая" версия платформы Flash Lite, чья функциональность ограничена в расчёте на возможности мобильных операционных систем и их аппаратных показателей.

В качестве основных средств разработки используются проприетарные пакеты Adobe Flash Professional и Adobe Flash Builder 4 (бывш. Adobe Flex Builder), позволяющие создавать интерактивные приложения (в том числе, веб-приложения, игры и мультфильмы).

Стандартным расширением для скомпилированных Flash-файлов (анимации, игр и интерактивных приложений) является .SWF (Shockwave Flash). Видеоролики в формате Flash представляют собой файлы с расширением FLV (при этом Flash в данном случае используется только как контейнер для видеозаписи). Расширение FLA соответствует формату рабочих файлов в среде разработки.

В основе Flash лежит векторный морфинг, то есть плавное «перетекание» одного ключевого кадра в другой. Это позволяет делать сложные мультипликационные сцены, задавая лишь несколько ключевых кадров для каждого персонажа.

Flash использует язык программирования ActionScript, основанный на ECMAScript.

ТЕХНОЛОГИИ РАСПОЗНАНИЯ ТЕКСТОВ В SWF ФАЙЛАХ

Для распознания текстов в SWF файлах необходимо знать структуру файлов этого типа. Данные файлы являются результатом работы компилятора и упаковщика, следовательно для выделения каких-либо объектов из таких файлов необходимо произвести декомпиляцию либо распаковку требуемого участка кода SWF. Технология декомпиляции файлов, содержащих коды для ядра Flash основывается на операции, обратной компилированию Action Script. То есть на выходе декомпилятора мы получим файл с текстом на языке Action Script, при компилировании которого будет получен точно такой же SWF файл. Однако, результат работы декомпилятора будет отличаться от исходного кода.

Полная распаковка с декомпилированием SWF файла средней сложности может занять относительно долгое время, поэтому стоит проблема декомпиляции и распаковки отдельных участков данного файла.

АНАЛОГИЧНЫЕ РЕШЕНИЯ

Аналогичные разрабатываемому алгоритмы в настоящее время используются в программах-декомпиляторах, прораммах для распознавания содержимого SWF файлов и в редакторах, которые служат для получения, замены и редактирования исходных данных SWF файла таких как изображения, аудио-файлы, шрифты, тексты и исходный код Action Script.

Алгоритмы распаковки SWF файлов используются в таких программах, как Sothink SWF Decompiler, SWF Decompile Expert, DecompileFlash_Free, Flash Decompiler Trillix, SWF Catcher.

Распознание текстового содержимого SWF файлов также производят поисковые роботы таких систем как “Google” и “Yahoo!”.

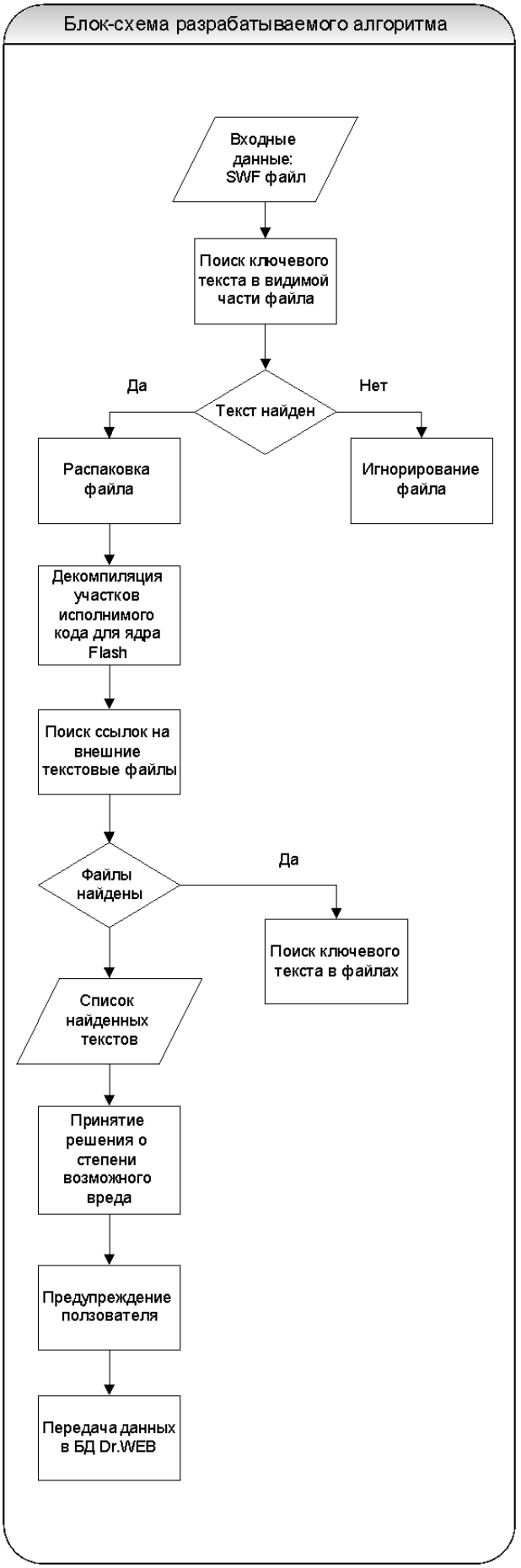
Алгоритмов, направленных на распознание вредоносных флэш-сайтов, содержащих просьбу об отправлении SMS на короткий номер, нет.

ПРЕДЛАГАЕМОЕ РЕШЕНИЕ

Для распознания вредоносных флэш-сайтов, содержащих просьбу об отправлении SMS используется разработанный алгоритм, который:

- производит поиск ключевого текста в видимой пользователю части файла SWF и принимает решение о дальнейшем анализе файла;
- производит просмотр всего SWF файла, пропуская выделение больших и ненужных в данной работе объектов, таких как изображения, аудио-файлы;
- производит декомпиляцию кодов Action Script с целью поиска в них ссылок на внешние XML файлы, содержащие описание текстовых объектов;

- проверяет найденные внешние текстовые файлы, используемые SWF файлом на наличие ключевого текста;
- производит выделение существующих текстовых объектов из SWF файла и проверяет их на наличие ключевого текста;
- принимает решение о степени возможного вреда от данного сайта;
- предупреждает пользователя о возможной угрозе;
- передаёт URL сайта и другую информацию о нём в базу данных Dr.WEB.



ОСОБЕННОСТИ ПРЕДЛАГАЕМОГО РЕШЕНИЯ

Разрабатываемый алгоритм описывается следующими особенностями:

- обладает возможностью распаковывать файлы SWF и получать из них необходимые данные;
- производит декомпиляцию участков файла SWF, являющихся кодами для ядра Flashж
- производит поиск ключевого текста, который является признаком вредоносных сайтов, содержащих просьбу отправить SMS;

Достоинствами данного алгоритма по сравнению с существующими алгоритмами чтения и распознания содержимого SWF файлов являются:

- производит принятие решения о степени возможной опасности флэш-сайта для посетителей;
- производит проверку внешних файлов на наличие ключевого текста;
- предупреждает посетителей о возможной опасности и о степени данной опасности;
- предаёт ссылку на вредоносный сайт в базу данных Dr.WEB.

Данный алгоритм может найти применение в следующих областях:

- для расширения возможностей антивирусных программ, фильтрующих сетевой поток между клиентом и сетью;
- для увеличения способностей поисковых роботов систем поиска с целью отсеивать вредоносные сайты при индексации;
- для пополнения антивирусных баз данных ссылками и информацией о вредоносных сайтах.

ЗАКЛЮЧЕНИЕ

Разрабатываемый алгоритм позволяет решить ряд недавно появившихся проблем в области информационной безопасности сети Интернет.

В дальнейшем развитии проекта предполагается:

- улучшить производительность алгоритма;
- улучшить алгоритм принятия решения;
- исследовать другие возможные реализации алгоритма распознавания текстов в SWF файлах;
- исследовать существующие программные средства, использующие подобные алгоритмы;
- включить возможность распознания текстов в изображениях.

ЛИТЕРАТУРА

1. Дмитрий Гурский. ActionScript 2.0: программирование во Flash MX 2004. Для профессионалов.
2. Герб Шилдт. Си для профессиональных программистов.
3. http://www.swftools.org – готовое решение распаковки и декомпиляции SWF файлов.