УДК 004.056

ПОИСК ВРЕДОНОСНЫХ ПРОГРАММ НА ОСНОВЕ АНАЛИЗА ПРОЦЕССА РАСПРОСТРАНЕНИЯ

И.А. Зикратов, Р.С. Василенко

Публикуется в порядке дискуссии

Определены основные проблемы традиционных методов обнаружения вредоносного программного обеспечения, основанных на обновлении антивирусных баз. Рассмотрены альтернативные методы, основанные на облачных вычислениях. Предложен новый метод обнаружения на основе анализа процесса распространения неизвестного программного обеспечения.

Ключевые слова: вредоносные программы, процесс распространения, репутационные сервисы.

Введение

Из отчетов ведущих антивирусных компаний за 2010–2011 г.г. [1–4] следует, что, помимо увеличения общего количества вредоносных программ, постоянно увеличиваются темпы появления нового вредоносного программного обеспечения (ПО). Практически все существующие технологии антивирусных продуктов, так или иначе, используют антивирусные базы на стороне пользователя, отсюда возникает проблема своевременного выпуска обновлений пользователям. Деятельность антивирусных компаний по выпуску обновлений антивирусных баз можно условно разделить на следующие этапы [5]:

- поступление образца в антивирусную лабораторию;
- анализ образца (ручной или, что чаще, автоматический);
- создание обнаруживающей записи (эвристической или по бинарным маскам);
- тестирование записи;
- выпуск баз обновлений.

Алгоритмы, используемые антивирусными компаниями для решения данной задачи, являются закрытыми и, вероятно, разными для каждой компании, однако есть один общий факт – каждый этап занимает определенное время. В среднем с момента попадания образца в антивирусную лабораторию до выхода обновления проходит время t, которое обычно не меньше двух часов. С начала распространения вредоносной программы до момента ее обнаружения проходит от 5 до 98 часов [5, 6].

Обладая данной информацией, создатели вредоносных программ оптимизируют свои алгоритмы выпуска вредоносных программ таким образом, чтобы максимально понизить эффективность выпущенных обновлений. В худшем случае это приводит к тому, что выпущенная антивирусной компанией обнаруживающая запись оказывается бесполезной, так как вредоносный образец, который она обнаруживает, уже прекратил свое существование.

Еще одной серьезной проблемой существующих методов обнаружения является требование наличия вредоносного файла в антивирусной лаборатории. Таким образом, с увеличением количества вредоносных программ антивирусным компаниям приходится обрабатывать все большие объемы файлов. По этой причине увеличивается вероятность как пропуска вредоносных объектов (в таком случае с большой вероятностью они вообще никогда не будут обнаружены), так и появления ложных срабатываний.

По оценкам Лаборатории Касперского, в 2010 г. было выявлено около 13 млн новых вредоносных программ. Если взять средний размер вредоносной программы порядка 1 МБ, то получится, что за 2010 г. лабораторией было обработано не менее 13 ТБ только вредоносных программ. При этом весь обработанный поток может быть в сотни раз больше. Своевременно скачать и обработать такой объем информации – серьезная проблема.

Описанные выше проблемы обусловливают необходимость разработки и реализации такого алгоритма обнаружения, который позволит сократить время каждого этапа. При этом целесообразно уменьшить зависимость эффективности алгоритма от обновлений антивирусных баз на стороне пользователя.

Целью данной работы является разработка нового метода обнаружения вредоносного ПО без использования антивирусных баз, основанного на анализе процесса распространения неизвестного ПО.

Репутационные сервисы

Стоит заметить, что с распространением широкополосного интернета, а также с увеличением количества пользователей антивирусных продуктов появляются новые возможности по сбору и анализу информации о появлении новых вредоносных программ, которые заключаются в том, что антивирусные компании получили возможность постоянно получать данные от пользователей, находящихся онлайн.

Предполагается, что эти изменения могут позволить разработать новый класс алгоритмов антивирусной защиты, основанных на анализе собираемой в реальном времени информации с машин пользователей. При этом принятие решений будет осуществляться на стороне серверов антивирусной компании, и за приемлемое время попадать к пользователям по сети.

Таким образом, можно сказать, что задача разработки класса гибких и адаптивных алгоритмов обнаружения новых вредоносных программ, не зависящих от процесса обновления антивирусных баз на стороне пользователя, является актуальной.

В работах [7–10] описаны некоторые подходы к решению сформулированной задачи. Основная идея, представленная в источниках, заключается в том, что, получая потоки статистических данных, а также метаданных [11] по данному объекту, они сопоставляют их имеющемуся набору логических правил отнесения объекта к чистому или вредоносному. Если правила срабатывают и при этом не противоречат друг другу, объект относится к одному из типов. Таким образом, в данных системах анализируются лишь количественные показатели, актуальные в определенный момент времени.

Процессы распространения программного обеспечения

Предлагаемый в работе подход отличается от известных учетом не только количественных показателей процесса распространения, но и использованием обоснованных качественных показателей этого процесса. Сутью разрабатываемого авторами подхода является предположение о том, что способы распространения вредоносных и легальных программ отличаются друг от друга. В отличие от легальных программ, которые пользователь скачивает «по собственному желанию», для распространения вредоносных программ их создателям приходится прибегать к множеству различных методов: эксплуатация уязвимостей, спам с вредоносными ссылками, социальная инженерия, против которых постоянно ведется противодействие со сторон антивирусных компаний, провайдеров и т.д. Это отражается на таких параметрах системы, как относительная скорость распространения (производная от функции распространения), а также скорость изменения относительной скорости распространения – «ускорение» (вторая производная от функции распространения).

Для проверки обоснованности этого предположения следует выполнить:

- выбор модели процесса распространения чистого и вредоносного ПО;
- определение параметров модели, т.е. выбор пространства признаков, описывающих процессы распространения;
- сбор информации, составление репрезентативной выборки распространения чистых и вредоносных программ;
- проведение экспериментов;
- анализ результатов и выбор наилучшего решения.

В качестве модели предполагается использовать функцию, описывающую процесс распространения чистого и вредоносного ПО во времени. Параметрами (пространством признаков) этой функции являются количество запусков и скачиваний ПО, а также количество уникальных пользователей.

Появление широкополосного Интернета у большинства пользователей антивирусных программ позволяют получать такие данные практически в реальном времени от всех пользователей сразу. Таким

образом, появляется возможность отслеживать возникновение, а также распространение новых неизвестных программ. В процессе распространения программы меняются и выбранные признаки. Предполагается, что на основе анализа характера и частоты изменений признаков, и сравнения полученных данных с базой данных процессов распространения известных программ можно отнести исследуемые образцы ПО к тому или иному типу. В отличие от всех существующих методов, в предложенном авторами подходе предлагается использовать не количественные показатели на какой-то определенный момент времени, а качественные изменения показателей во времени.

На рис. 1, 2 представлены полученные экспериментально зависимости параметров распространенности вредоносного и чистого ПО от времени. Эксперимент проводился на статистических данных, полученных от пользователей Лаборатории Касперского. В ходе эксперимента отслеживались параметры функций распространения неизвестного ПО (количество запусков). Объекты, которые впоследствии признавались антивирусом вредоносными, попадали в группу вредоносного ПО. Объекты, признанные в итоге чистыми, попадали в группу чистого ПО. После этого анализировались характеристики функции распространения для той и другой группы. На рисунках приведены самые характерные функции распространения для чистого и вредоносного ПО.

Из вредоносных программ были взяты:

- Trojan-Ransom.Win32.DigiPog;
- Trojan-PSW.Win32.Dybalom;
- Trojan.Win32.VBKrypt;
- Trojan.Win32.Inject.

Из чистых программ:

- Adobe Shockwave Player;
- eRightSoft SUPER setup;
- CDBurnerXP;
- Internet Download Manager installer.

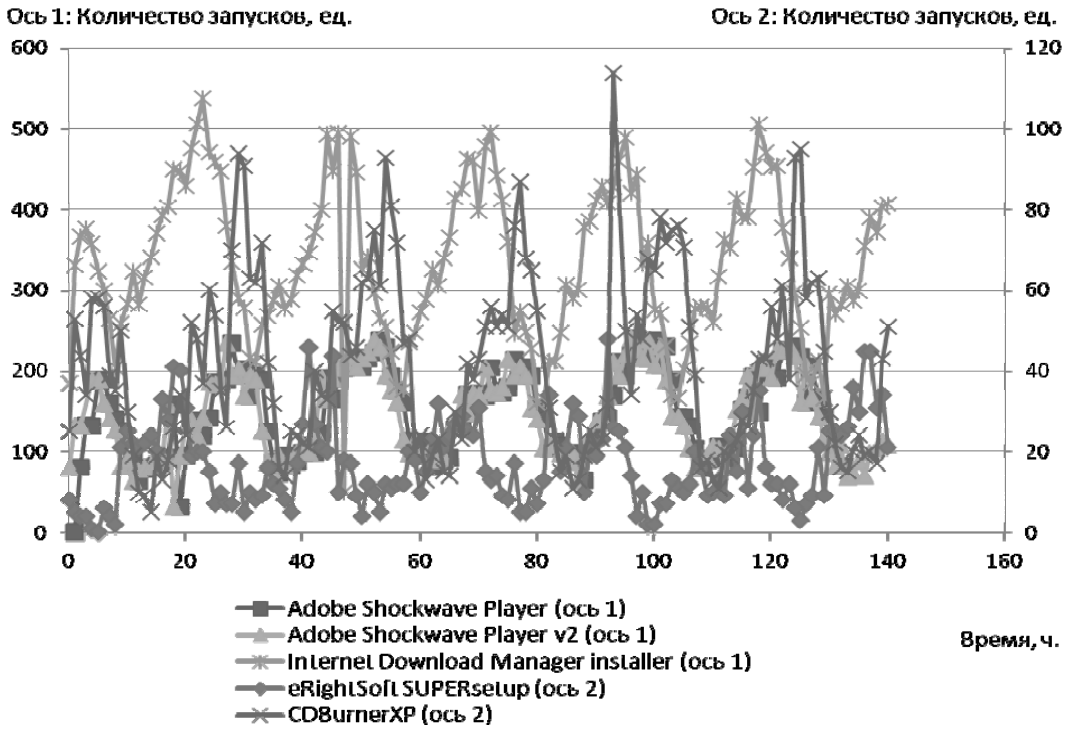

Рис. 1. Функции, описывающие процесс распространения чистого ПО

Как видно из рисунков, параметры кривых существенно отличаются. Для чистого ПО (рис. 1) характерны:

- длительное время жизни;
- устойчивость циклов на суточных промежутках;
- плавные, обусловленные изменением макропараметров (таких, как время суток или выход новой версии), изменения скорости распространения.

В то же время как для вредоносного ПО (рис. 2) характерны:

- короткое время жизни;

– отсутствие четкой зависимости от суточного периода;
– отсутствие плавного затухания скорости распространения – характерна резкая остановка распространения;
– «скачкообразные» изменения для скорости распространения.

Отличия обусловлены тем, что на процесс распространения вредоносной программы влияет намного большее количество факторов, нежели на процесс распространения легального приложения, одни из которых способствуют увеличению скорости распространения, другие – ее уменьшению.

Таким образом, сформулированная ранее задача сводится к задаче распознавания образов в новом признаковом пространстве. Очевидно, что воздействие большого количества стохастических факторов на процессы распространения ПО, вредоносного и невредоносного, повлечет флуктуации кривых на рис. 1, 2. Это обусловливает возможность использования статистических алгоритмов для решения задачи распознавания образов.

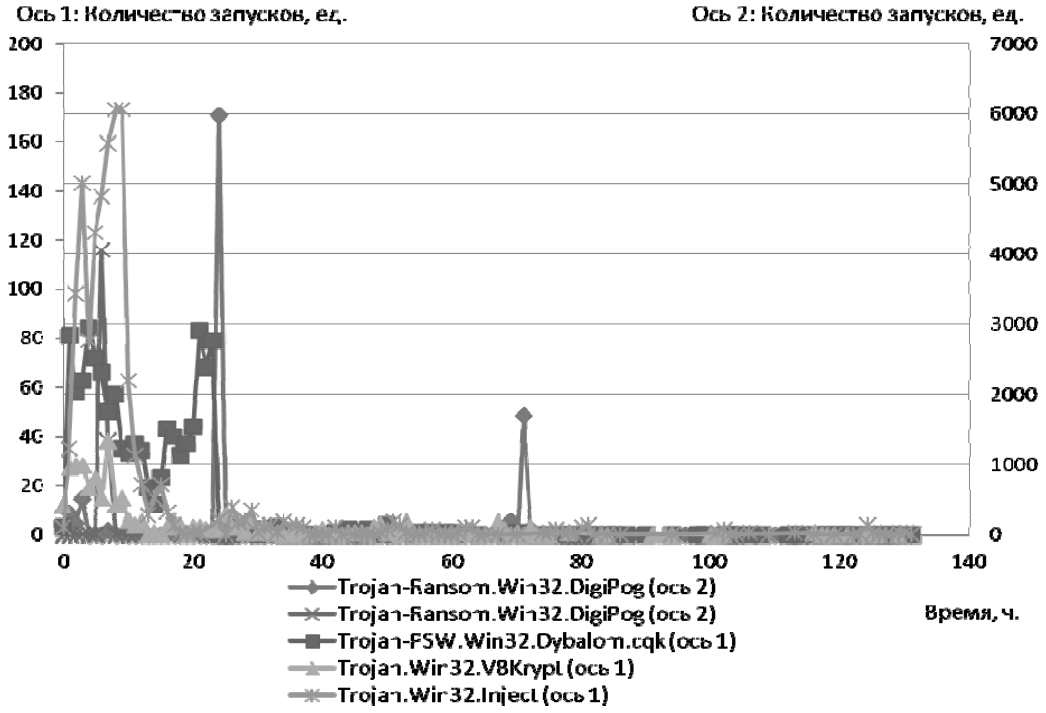

Рис. 2. Функции, описывающие процесс распространения вредоносного ПО

Определившись с пространством признаков, далее требуется сформировать базу знаний, например, в виде линейных матриц, как о процессах распространения заведомо легальных программ, так и о процессах распространения вредоносных программ. В ходе создания базы знаний, а также проведения испытаний признаки подобия процессов-паттернов могут уточняться введением дополнительных корреляций между первичными параметрами.

Целью дальнейших исследований является обоснование и разработка классификатора, который позволит с заданной степенью достоверности осуществлять задачу классификации образов.

Заключение

Предложен новый подход к обнаружению вредоносных программ, основанный на анализе процесса их распространения, отличающийся от известных тем, что рассматриваются не отдельные признаки, описывающие содержание файла, а параметры, характеризующие то, как файл распространялся у пользователей.

На основе данного подхода может быть разработан новый класс алгоритмов, основанных на собираемой в реальном времени информации о распространении некого программного обеспечения и классифицирующих данное программное обеспечение как чистое или вредоносное. Планируется, что новые алгоритмы будут опираться на выдвинутое в работе предположение, согласно которому процессы распространения вредоносного и легального программного обеспечения различаются.

Также несомненным достоинством предложенного метода является возможность предотвращения ложных срабатываний на тех файлах, процесс распространения которых похож на процесс распространения чистых программ.

Для дальнейшей работы над представленным методом целесообразно описать и оценить:

- выбор параметров модели (алгоритм принятия решения, задачи классификации образов);
- информативность признакового пространства;
- эксперименты и анализ их результатов.

Литература

1. Гостев А. Kaspersky Security Bulletin 2010. Развитие угроз в 2010 году [Электронный ресурс]. – Режим доступа: http://www.securelist.com/ru/analysis/208050677/Kaspersky_Security_Bulletin_2010_Razvitie _ugroz_v_2010_godu, свободный. Яз. рус. (дата обращения 29.11.2011).
2. Наместников Ю. Развитие информационных угроз во втором квартале 2011 года [Электронный ресурс]. – Режим доступа: http://www.securelist.com/ru/analysis/208050710/Razvitie_informatsionnykh _ugroz_vo_vtorom_kvartale_2011_goda, свободный. Яз. рус. (дата обращения 29.11.2011).
3. Наместников Ю. Развитие информационных угроз в первом квартале 2011 [Электронный ресурс]. – Режим – доступа: http://www.securelist.com/ru/analysis/208050695/Razvitie_informatsionnykh_ugroz_v_ pervom_kvartale_2011_goda, свободный. Яз. рус. (дата обращения 29.11.2011).
4. Fossi M., Mazurek D., Egan G. Symantec Internet Security Threat Report 2010 [Электронный ресурс]. – Режим доступа: https://www4.symantec.com/mktginfo/downloads/21182883_GA_REPORT_ISTR_Main-Report_04-11_HI-RES.pdf, свободный. Яз. англ. (дата обращения 29.11.2011).
5. Машевский Ю. Антивирусный прогноз погоды: облачно [Электронный ресурс]. – Режим доступа: http://www.securelist.com/ru/analysis/208050657/Antivirusnyy_prognoz_pogody_oblachno, свободный. Яз. рус. (дата обращения 29.11.2011).
6. Nachenberg C., Ramzan Z., Seshadri V.Reputation: a new chapter in malware protection // Virus Bulleti Conference. – 2009. – P. 185–191.
7. Elovici Y.M., Tachan G.O., Shabtai A.C. A system that provides early detection, alert, and response to electronic threats. – European patent app. No. 07015353.1, 2008.
8. Mashevsky Y.V., Namestnikov Y.V., Denishchenko N.V. Detection and minimization of false positives in anti-malware processing. – US Patent No. 7,640,589 B1, 2009.
9. Mashevsky Y.V., Namestnikov Y.V., Denishchenko N.V. Method and system for detection and prediction of computer virus-related epidemics. – US Patent No. 7,743,419 B1, 2010.
10. Judge P. System and method for message threat management. – US Patent No. 7,225,466 B2, 2007.
11. Mering M., Childers S., Fleming A. Report of Task Force on Metadata Analysis // American Library Association. – Lincoln, 2006. – 17 p.

Зикратов Игорь Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, доцент, зав. кафедрой, zikratov@cit.ifmo.ru

Василенко Роман Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, zyx2145@gmail.com