ПОЛИМОРФНЫЕ ПРЕОБРАЗОВАНИЯ ВРЕДОНОСНОГО КОДА И ВОЗМОЖНЫЕ ПУТИ ИХ ОБНАРУЖЕНИЯ

Дроботун Е. Б.

Военная академия воздушно-космической обороны, г.Тверь

201074@mail.ru

В работе рассмотрены основные принципы генерации полиморфного кода вредоносного программного обеспечения, препятствующие его обнаружению антивирусными программами, а также рассмотрены возможные методы обнаружения и анализа таких преобразований. В дополнение к существующим методам предлагается метод анализа полиморфных преобразований, основанный на анализе частоты появления отдельных команд и подсчете энтропии исследуемого кода.

Термин «полиморфизм» (с греч. «многоформенность») применительно к компьютерным вирусам появился приблизительно в 1990 году. С того времени полиморфизм в вирусах прошел множество стадий своего развития: от простейшего побайтного шифрования до преобразований кода, использующих сложнейшие алгоритмы, в том числе криптографические.

Анализ работы большинства антивирусных программ показывает, что для гарантированного обнаружения компонентов вирусного или вредоносного кода используется сигнатурный поиск с жесткой привязкой к точке входа или физическому смещению в файле. Сигнатурой называется уникальная для каждого вируса последовательность байтов, однозначно его идентифицирующая. Сигнатура может быть сплошной (к примеру, DEh ADh FDh 3Eh F2h), либо разреженной (например, DEh … EDh … 3Ah C2h … EFh). Разреженная сигнатура иначе называется маской, поиск по разреженной сигнатуре – поиском по маске [1].

Для достижения приемлемой скорости сканирования антивирусные программы при сигнатурном поиске практически никогда не проверяют весь файл целиком, а ограничиваются лишь проверкой нескольких ключевых точек (например, в точке входа либо с привязкой сигнатуры к физическому смещению относительно начала файла).

Полиморфные вирусы – это достаточно трудно обнаруживаемые вирусы, не имеющие сигнатур, то есть, не содержащие ни одного постоянного участка кода. В большинстве случаев два образца одного и того же полиморфного вируса не будут иметь ни одного совпадения. Это достигается шифрованием основного тела вируса и модификациями программы-расшифровщика (декриптора). К полиморфным вирусам относятся те из них, детектирование которых невозможно (или крайне затруднительно) осуществить при помощи так называемых вирусных масок - участков постоянного кода, специфичных для конкретного вируса. Достигается это двумя основными способами - шифрованием основного кода вируса с непостоянным ключом и случайным набором команд расшифровщика или изменением самого выполняемого кода вируса.

Первым представителем полиморфных вирусов стал "Chameleon" в начале 1990 года, однако проблема стала серьезней чуть позже – в апреле 1991 года была зарегистрирована эпидемия "Tequila". Идея полиморфных вирусов стала столь популярна, что дошло до создания генераторов полиморфных вирусных кодов – первым стал MtE. Кроме того, генератор позволял получать полиморфный вирус из обычного – путем присоединения к OBJ-файлу вируса файла полиморфного кода с идентичным расширением [2].

Фактически, с появлением генераторов для создания полиморфного вируса не требовалось знать код оригинального вируса – достаточно было просто "скормить" его генератору, и тот делал всю работу за начинающего хакера. Впоследствии полиморфные вирусы стали крайне популярны, так как для их отлова требуются специальные математические алгоритмы восстановления исходного кода вируса, эмуляция исполняемого вирусом действий и другие сложности.

Генераторы полиморфных вирусов также совершенствовались – в середине девяностых это были MTE 0.90 (Mutation Engine), TPE (Trident Polymorphic Engine), четыре версии NED (Nuke Encryption Device) и DAME (Dark Angel's Multiple Encryptor).

Полиморфные вирусы активно развивались примерно до начала XXI века, однако затем произошел общий крен вирусописательства в сторону червей и троянов. Технология постоянной мутации кода для затруднения обнаружения антивирусными программами временно оказалась невостребованной.

Однако, начиная примерно с 2003 года, полиморфизм снова начинает привлекать внимание вирусного сообщества. Это было вызвано тем, что антивирусные программы все больше и больше совершенствовались, и уже стало нельзя использовать в качестве инструментов скрытия кода различные программы-паковщики, которые были на тот момент излюбленным детищем вирусописателей.

В настоящее время практически все вирусы используют технику полиморфизма в том или ином виде для скрытия себя от антивирусных программ [2]. В зависимости от сложности алгоритмов полиморфного преобразования, используемых вирусами, можно выделить несколько уровней полиморфизма:

Уровень 1. Вирусы первого уровня полиморфизма используют постоянные значения для разных расшифровщиков. Их можно обнаружить по некоторым постоянным участкам кода расшифровщика. Такие вирусы принято называть "не совсем полиморфными", или олигоморфными (oligomorphic).

Уровень 2. Ко второму уровню полиморфизма относят вирусы, расшифровщик которых имеет постоянной одну или несколько инструкций. Например, он может использовать различные регистры, некоторые альтернативные инструкции в расшифровщике. Такие вирусы также можно распознать по определенной сигнатуре - заданным сочетаниям байт в расшифровщике.

Уровень 3. Вирусы, использующие в расшифровщике команды, не участвующие в расшифровании вирусного кода, или "команды-мусора", относят к третьему уровню полиморфизма. Это такие команды ассемблера, как NOP, MOV AX, AX, STI, CLD, CLI и т.д. Данные вирусы также можно определить с помощью некоторой сигнатуры, если произвести отсеивание всех "мусорных" команд.

Уровень 4. Вирусы четвертого уровня используют в расшифровщике взаимозаменяемые инструкции и "перемешанные" инструкции без изменения алгоритма расшифрования. Например, ассемблерная команда MOV AX,BX имеет взаимозаменяемые инструкции: PUSH BX - POP AX; XCHG AX,BX; MOV CX,BX - MOV AX,CX и т. д. Детектирование данных вирусов возможно с помощью некоторой перебираемой сигнатуры.

Уровень 5. Пятый уровень полиморфизма включает свойства всех перечисленных уровней, а кроме того, расшифровщик может использовать различные алгоритмы расшифрования вирусного кода. Для расшифровки возможно использование основного вирусного кода, расшифровки части самого же декриптора или нескольких расшифровщиков, поочередно расшифровывающих друг друга либо непосредственно вирусный код. Как правило, обнаружение вирусов данного уровня полиморфизма с помощью сигнатуры невозможно. Если для обнаружения такого вируса возможен серьезный анализ кода только самого расшифровщика, то для лечения необходимо произвести частичную или полную расшифровку тела вируса, чтобы извлечь информацию о зараженном файле.

Уровень 6. К нему относятся нешифрованные вирусы - т. е. вирусы, состоящие из программных единиц-частей, которые "перемешиваются" внутри тела вируса. Данные вирусы, как "кубики", тасуют свои подпрограммы (инсталляции, заражения, обработчика прерывания, анализа файла и т. д.). Такие вирусы еще называются пермутирующими (permutating).

Если для поиска и обнаружения вирусов от первого до четвертого уровня полиморфизма можно использовать более усовершенствованные алгоритмы сигнатурного поиска, вирусы пятого и шестого уровня полиморфизма, не содержащие ни одной постоянной последовательности байтов, с помощью сигнатурного поиска не обнаруживаются и для их детектирования используется техника эмуляции процессора (называемая также технологией виртуальной машины). Антивирусная программа выполняет код в эмуляторе и после того как декриптор расшифрует основное тело вируса далее применяет сигнатурный поиск. Это очень ресурсоемкая операция, и если применять ее для каждого исследуемого файла, процесс проверки будет занимать очень много времени.

Для сокращения времени проверки файлов, можно эмулировать выполнение исследуемых файлов, только для файлов в которых предположительно обнаружены полиморфные прнеобразования.

Литература

1. *Касперски К.* // Компьютерные вирусы изнутри и снаружи. – СПб.: Питер, 2006. – 527 с.: ил.
2. *Касперски К.* // Записки исследователя компьютерных вирусов. – СПб.: Питер, 2005. – 316 с.: ил.