РЕКОНСТРУКЦИЯ ЗАКРЫТЫХ ФОРМАТОВ ПОЧТОВЫХ БАЗ (MS OUTLOOK, MS OUTLOOK EXPRESS)

С.И. Уласень;
Г.К. Резников; к.т.н., с.н.с.
ОДО «ВирусБлокАда»

Почтовая база представляет собой сложный контейнерный объект, который может содержать внутри себя вредоносные вложения. Перед антивирусом встает задача проверки почтовых баз на наличие в них вредоносных объектов и корректного их лечения. Какие задачи должны поставить перед собой разработчики антивируса при поиске вредоносных файлов в почтовых базах? С какими трудностями они сталкиваются?

Зачем это нужно. Почему эта проблема актуальна. Пути попадания вредоносных объектов к пользователю через почтовые протоколы

Первым (и самым массовым) фактором попадания вредоносных объектов к пользователю через почтовые протоколы является спам-рассылка. Доля спама в почтовом трафике в этом году составляет порядка 85%. При этом примерно 1% этих писем содержит в себе вредоносные вложения. Зачастую такие письма либо сразу же удаляются самим пользователем, либо срабатывают методы социальной инженерии, после чего пользователь открывает вредоносное вложение, а уже после удаляет письмо из ящика. Пользователь очень редко станет хранить спам.

Вторым (и самым опасным) путем попадания является пересылка файлов между пользователями. Пользователи обмениваются зараженными файлами между собой, сами того не подозревая. Пользователь, получивший такое письмо, зная, что письмо получено из достоверного источника, тут же им воспользуется. Если на момент получения вообще не установлен антивирус, либо не используется система почтового фильтра или антивирусный плагин к почтовой системе, произойдет заражение каким-нибудь вирусом, либо внедрение трояна или червяка. Кроме того, в данном случае письмо не будет удалено пользователем из ящика, а останется там для создания истории переписки или в надежде его использования еще раз в будущем. И именно здесь таится самая настоящая опасность.

Многие пользователи, особенно в корпоративной среде, работают с почтовыми клиентами. Попадая в базу почтового клиента, вышеупомянутое письмо может храниться там весьма и весьма долго. А это уже приводит к тому, что пользователь не только не замечает, что к нему попало такое «опасное» письмо, но и хранит его.

Еще одной очень важной особенностью почтовых баз является то, что они имеют обыкновение мигрировать вместе с пользователем. Так, если пользователь переустанавливает операционную систему, он «бэкапит» почтовую базу и восстанавливает работу с ней после. Если пользователь переходит работать на другой компьютер, он также забирает с собой и почтовую базу. После чего он дальше продолжает ей пользоваться. В какой-то момент ему может понадобиться информация из ранее полученного письма, после этого он находит его в базе и начинает запускать прикрепленные файлы. Ситуация опять повторяется.

Но представим себе ситуацию, что в какой-то момент на компьютере пользователя появился антивирус. Перед антивирусом тут же встает задача обеспечения безопасной работы пользователя с теми данными, которые уже есть у него в почтовой базе. Далее рассмотрим те методы, которые должен предоставить антивирус пользователю для решения данной задачи.

Какие задачи должны поставить разработчики антивируса перед собой при решении задачи безопасной работы пользователя с теми данными, которые есть в почтовой базе

Самый первый путь, по которому могут пойти разработчики, это реализовать модуль-плагин к почтовому клиенту. Данный модуль должен иметь возможность не дать пользователю открыть вредоносные вложения, а также проверить вложения получаемых писем. Однако, к сожалению, данный модуль не решит всех тех задач, которые ставятся перед антивирусом. Потому необходимы и иные решения.

Представим себе ситуацию, что где-то на компьютере пользователя или на файловом сервере хранится «бэкап» почтовой базы. Пользователь хочет просканировать антивирусным сканером директорию, в которой расположен данный «бэкап», для того, чтобы убедиться, что в ней нет вредоносных объектов. Или же, что часто бывает, «бэкап» почтовой базы пересылается при помощи какого-либо протокола между пользователями. При этом на каком-то этапе данный объект также проверяется антивирусом. В обоих случаях антивирус должен вынести однозначный вердикт о том, является ли данный объект инфицированным, либо содержит в себе инфицированные объекты.

Исходя из того, что почтовая база представляет собой сложный контейнерный объект, перед разработчиками возникают следующие задачи:

1. Поиск и проверка вложений на наличие вредоносных объектов;
2. Вывод результатов проверки с указанием точного местоположения проверяемых объектов в базе;
3. Нахождение способа корректного лечения/удаления вредоносных вложений в базе.

В мире разработано не так уж и много почтовых клиентских систем. А действительно популярных среди них еще меньше. Одними из таких популярных приложений являются почтовые клиентские системы от компании Microsoft. Это системы MS Outlook и MS Outlook Express. Далее поговорим о них.

С какими трудностями столкнутся разработчики

Первая трудность, с которой столкнется разработчик, - это невозможность получения из открытых источников официальной документации по формату баз. При этом Microsoft предоставляет API для доступа к базам. Но воспользоваться данным API не предоставляется возможным. Во-первых, данные библиотеки не являются кросс-платформенными, и ими невозможно воспользоваться для проверки файла базы, например, под Linux. Во-вторых, даже под Windows необходимо, чтобы на компьютере, на котором происходит проверка, была установлена почтовая система. Либо данные библиотеки необходимо брать в состав антивирусного ядра, что неприемлемо ни с точки зрения лицензий, ни объема антивирусного ядра. Потому данный путь тупиковый. Единственное, что остается разработчику – это заняться реконструкцией данного формата.

И тут еще на этапе подготовки выяснится, что в формат базы PST в MS Outlook периодически вносятся какие-то незначительные изменения. А перед выходом MS Outlook 2003 формат ее базы был значительно переработан и изменен. Вследствие чего номер версии формата базы изменился (до этого он не изменялся ни разу). Так размер внутренних полей был расширен для поддержки UNICODE и 64-битных ссылок. И сама база теперь может превышать размер в 4 Гб.

Сам формат PST имеет древовидную структуру. И по своему виду чем-то напоминает формат NTFS. Поэтому на этапе реконструкции у нас даже закрались подозрения, что к этим двум форматам приложили руку одни и те же разработчики. Так, например, «аттач», прикрепленный к письму, будет разбит в базе на несколько фрагментов определенного размера. При этом в листе дерева может быть расположена структура, в которой описан каждый из фрагментов. А если «аттач» небольшой по размеру, то он сам будет располагаться в данном листе.

Еще одной проблемой станет то, что MS Outlook позволяет шифровать данные, хранящиеся в его базе. Для этого база поддерживает два уровня шифрования: Compressible Encryption и High Encryption. Тут есть один подводный камень - это то, что в русской версии MS Outlook данные пункты переведены неправильно и подаются как сжатие данных.

Тут же встает дилемма, нужно ли антивирусу проверять те данные, которые являются шифрованными? Мы посчитали, что нужно.

После того, как вся эта работа будет проделана, и разработчики выполнят первые два требования (поиск и проверка объектов, вывод результатов), останется самое сложное – лечение в базе. А самое сложное (и ответственное) здесь то, что некорректное лечение базы может привести к полному разрушению ее формата и база будет потеряна. Мы с большой осторожностью отнеслись к данному пункту.

Как эти трудности героически победить

Быстро победить, к сожалению, не получится. Реконструкция закрытых форматов – процесс длительный и кропотливый.

Сначала мы взялись за MS Outlook Express. Для этого было решено воспользоваться проектами с открытым исходным кодом, в которых решаются задачи по конвертации базы в другой формат. Но это оказалось делом совершенно неблагодарным и ни к чему не привело. Было решено проделать всю работу самим с нуля. Для этого были созданы сотни баз, сразу же пустых, потом в них письма добавлялись, удалялись, в них изменялось содержимое. На каждом из этапов распечатывались дампы файлов, и шло их скрупулезное сравнение. В итоге времени было затрачено много, но поставленного результата мы добились.

Затем мы взялись за MS Outlook. На тот момент еще не вышла 2003 версия MS Office, и мы имели дело с первой версией формата PST.

Выяснилось, что в состав программы MS Outlook входит утилита scanpst.exe. Задача данной утилиты – восстановление битых почтовых баз PST. Но кроме всего прочего данная утилита имеет набор недокументированных параметров командной строки. Задавая недокументированные параметры и предоставляя утилите битую базу, можно получить примерное (не всегда точное) представление той структуры, в которой присутствуют неверные значения. Для этого также создавались сотни баз, делались распечатки их дампов.

В итоге оказалось, что формат PST намного сложнее и запутаннее формата DBX. Но разобраться с ним получилось быстрее. После того как вышел MS Outlook 2003, данную работу пришлось проделать заново, так как формат базы значительно изменился.

С проблемой шифрованных вложений справились также быстро. Потому что шифрование уровня Compressible Encryption подразумевает под собой обычное шифрование по алгоритму Цезаря. А уровень High Encryption хоть и сложнее, но особых трудностей тоже не составил.

Потом мы подобрались к лечению в базе.

Как уже описано выше, вложение в базе PST может быть разбито на фрагменты. Каждый из этих фрагментов подписан CRC. Потому если в базе мы обнаруживаем троян, то все его фрагменты мы затираем нулями с соответствующими подписями CRC. Но что, если в базе обнаружен файл, инфицированный каким-либо вирусом? Вот тут встает непростая задача. Лечение вируса зачастую ведет к изменению размера его образа. Соответственно лечение в базе может привести к изменению ее структуры. А это достаточно серьезный шаг, который при малейшей ошибке может привести к необратимым последствиям. Мы не стали рисковать. Точнее рискнули совсем чуть-чуть. Если лечение данного вируса ведет к изменению его размера, то мы затираем его нулями, если же не ведет – лечим.

На этом работа по обнаружению и лечению в базах MS Outlook и MS Outlook Express была закончена.

Таким образом, по результатам проведенной работы можно сделать следующие выводы:

1. Рассматриваемая проблема является актуальной, потому что именно через почтовые протоколы чаще всего осуществляется попадания вредоносных объектов к пользователю;
2. Модуль-плагин к почтовому клиенту не решает проблем, связанных со сканирование почтовой базы;
3. При сканировании почтовой базы необходимо решить следующие задачи:
   - поиск и проверка вложений на наличие вредоносных объектов;
   - вывод результатов проверки с указанием точного местоположения проверяемых объектов в базе;
   - нахождение способа корректного лечения/удаления вредоносных вложений в базе.
4. Почтовая база представляет собой сложный контейнерный объект, структура которого недостаточно описана и периодически изменяется;
5. Решение задач, возникающих при сканировании почтовой базы, невозможно без проведения экспериментальных работ.

В результате проведенных работ был создан специальный модуль, и в настоящее время комплекс Vba32 программных средств защиты от воздействия вредоносных программ имеет в своем составе мощное средство антивирусной обработки наиболее распространенных почтовых баз.