virustotal своим руками, лапами и хвостом

крис касперски ака мыщъх, a.k.a. nezumi, a.k.a. souriz, a.k.a. elraton, no-email

онлайновая антивирусная служба www.virustotal.com завоевала бешенную популярность. многие хотят прикрутить к своему сайту что-то похожее, но не знают как. зачем это нужно? во-первых, чтобы поднять посещаемость, а, во-вторых, пополнить собственную коллекцию вирусов, червей и прочей малвари свежими экземплярами. www.virustotal.com не раскрывает своих секретов, но мыщъх уже давно разобрал его по винтикам и теперь работает над улучшенной реализаций

введение

В то время как одни пользователи держат на компьютере целый зоопарк различных антивирусов, конфликтующих друг с другом и тормозящим ПК (не говоря уже о стоимости лицензий или сложности поиска правильного "лекарства"), хакеры предпочитают ловить малварь самостоятельно, в крайнем случае проверяя подозрительные файлы на бесплатных онлайновых службах типа того же www.virustotal.com. Эти же службы используются для "обкатки" вирусов собственного написания на предмет их детекции эвристическими анализаторами. И хотя, если верить блогу Евгения Касперского (www.viruslist.com/en/weblog), хакеры не доверяют virus-total'y, поскольку он передает подозрительные файлы антивирусным компаниям и вирусы начинают палиться еще на излете, эта точка зрения отражает лишь малую часть действительности. Да, действительно, профессиональные разработчики атакующих программ и rootkit'ов проверяют их на "вшивость" исключительно локальным способом на своих собственных машинах, предотвращая утечку информации, но… профессионалов единицы, к тому же экспериментируя с virus-total'ом, хакеры определяют _общие_ критерии ругательства антивирусов, выявляя последовательности машинных команд/вызовов API-функций, приводящих к срабатыванию эвристического анализатора. Однажды "обломав" антивирус, хакер может _многократно_ использовать найденный примем обхода эвристика. Достаточно посетить любые форумы, где обитают вирусописатели, чтобы убедиться, что они весьма неравнодушны к virus-total'y и активно используют его в своих целях.

А что если создать еще более качественный сервис? Ведь virus-total примитивен до ужаса — качество сканирования оставляет желать лучшего, не говоря уже о длинных "социалистических" очередях в которых проходится подолгу простаивать из-за частых перегрузок сервера (а все потому что, балансировка нагрузки и оптимизация изначально не предусматривались!)

На момент написания этих строк, мыщъх по заказу одной антивирусной компании (имя которой не вправе разглашать из-за NDA) руководит разработкой онлайнового сервиса, рассчитанного на "магистральную" загрузку и предоставляющего пользователям кучу всевозможных рычагов управления. Естественно, исходный код к статье не прилагается, да он и не нужен, главное — это концепт, плюс некоторые неочевидные тонкости, с которыми придется столкнуться при "промышленных" масштабах эксплуатации.

Естественно, все это требует широких сетевых каналов, мощных многопроцессорных систем и еще кучу всего. Словом, без солидных финансовых вложений тут никаких не обойтись. Однако, никто же не заставляет нас создавать сервис планетарного масштаба и если постараться, то можно вполне уложиться в бюджет ~$2k или даже менее того. Нам потребуется интернет-канал с безлимитным тарифом (т. е. только абонплата), чтобы злые люди не кинули нас на входящий трафик, который в данном случае будет весьма значительным. В качестве компьютера вполне подойдет машина с процессором Core2Duo и парой гигабайт оперативной памяти. О проблемах с лицензированием антивирусов мы погорим в одноименной врезке, а пока же отметим, что никаких особых программистских навыков нам не потребуется. Подойдет любой язык (Си, Perl. PHP) и минимальный опыт работы с CGI (пользуясь случаем, хочется порекомендовать библиотеку CGIC).

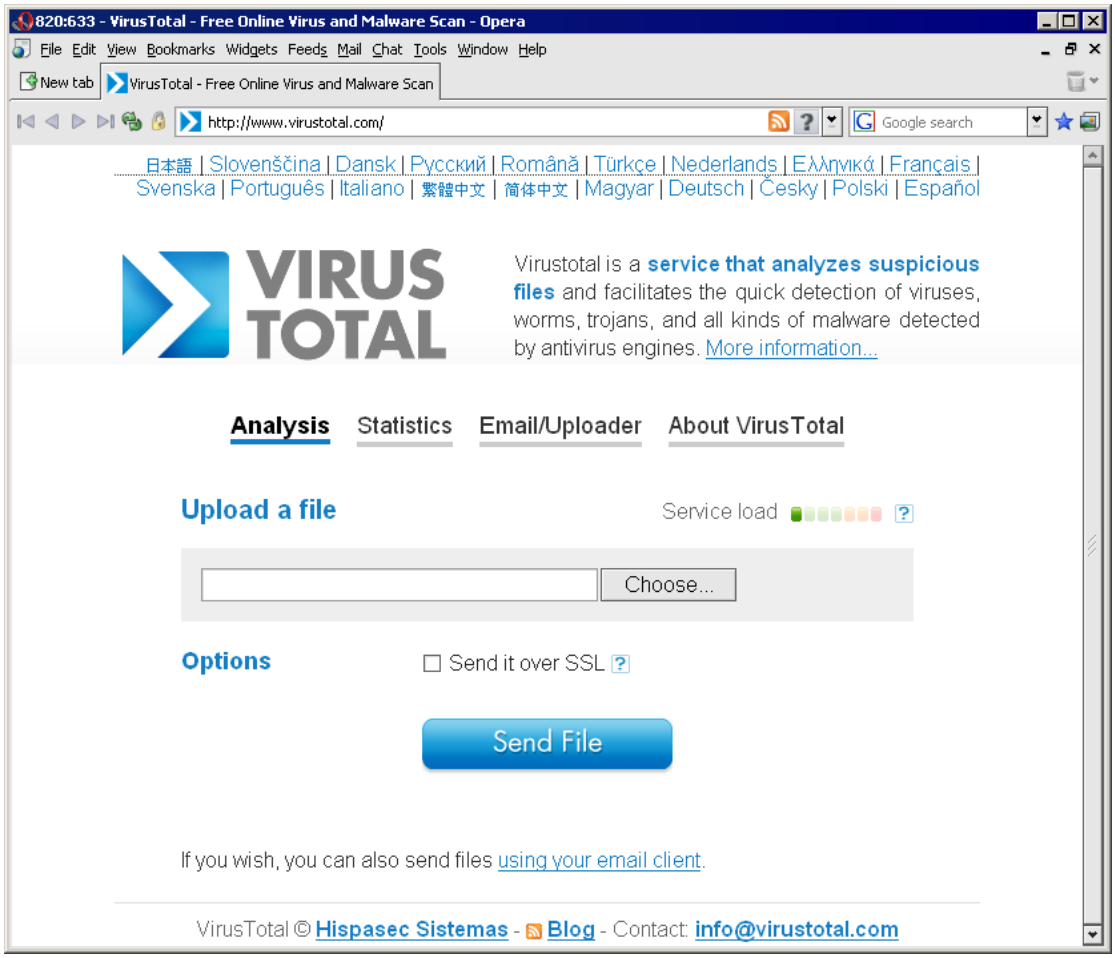

Рисунок 1 популярная онлайновая антивирусная служба virus-total – так она выглядит снаружи. а хотите узнать, что находится у нее под капотом?! часть информации можно найти на ее же собственном блоге, но… гораздо интереснее спроектировать улучшенную версию онлайнового антивирусного сканера с чистого листа

virus-total изнутри

Virus-total устроен не просто, а _очень_ просто. Он использует консольные версии антивирусов, управляемые посредством командой строки и выдающие результат сканирования в стандартный проток вывода, который легко перенаправить в файл или пайп (pipe).

Что мы делаем? Через специальную форму на сайте закачиваем "подопытный" файл, скармливаем его антивирусу, предварительно перенаправив вывод во временный файл/пайп, который тут же парсим (т. е. выбрасываем все лишнее, оставляя только статус проверки и имя вируса). Парсить вывод легче всего Perl'ом, поддерживающего мощный механизм регулярных выражений, однако, Си-программы намного более производительны, а потому более предпочтительны (особенно, при большом наплыве пользователей).

Собственно говоря, на этом возможности virus-total'а и заканчиваются, что создает большие проблемы: во-первых, далеко не все антивирусы имеют консольные версии, а, во-вторых, даже из тех, что имеют, их поведение зачастую радикально отличается от полноценных GUI-версий, в чем легко убедиться сравнив результаты сканирования большой коллекции вирусов локальным способом и через virus-total и это сравнение будет отнюдь не в пользу virus-total'а.

Учитывая, что практически все антивирусы (и GUI-версии в том числе) поддерживают запись результатов сканирования в log-файл и позволяют задавать имя сканируемого файла через командную строку или на худой конец через механизм DDE (Dynamic Data Exchange), ничего не стоит "прикрутить" GUI-версию к онлайновой службе. Просто "скармливаем" антивирусу файл, форсируем запись результатов сканирования в log-файл, который парсим так же, как и вывод консольных версий.

Остается только собрать "показания" всех имеющихся в нашем распоряжении антивирусов, оформить их в виде HTML-таблицы и выдать на экран, что по силам даже самым начинающим программистам.

В клинических случаях, когда антивирус начисто игнорирует командную строку или не умеет вести логи, на помощь приходит механизм Windows-сообщений (Windows Message или, сокращенно, WM). Посылая WM-сообщения элементам управления антивируса, мы можем манипулировать кнопками, меню и прочими элементами управления по своему усмотрению. Аналогичным способом извлекается и содержимое окна, содержащего результаты проверки. Получив форматированный rich-текст или plain-текст, пропускаем его через парсер и все!!!

Рисунок 2 онлайновая антивирусная служба в действии!

маленькие секреты больших серверов

При попытке практической реализации вышеописанной модели, неизбежно выплывут проблемы удручающее низкой производительности, требующие решения. Но мы не боимся трудностей и, покурив хорошей травы, начнем щемить проблемы одну за другой.

Первое и очевидное. Как показывает статистика, различные пользователи преимущество проверяют _одни_ и те же файлы, как правило, принадлежащие Windows или популярным программным пакетам. Чтобы сократить накладные расходы, рекомендуется подсчитывать контрольную сумму файла перед его проверкой и, если такой файл уже проверялся ранее, выдавать уже готовые результаты сканирования, сохраненные в базе данных. На первый взгляд, в реализации данного алгоритма нет ничего сложного, но тут притаилось немало подводных камней, вот только наиболее актуальные из них:

- антивирусные базы обновляются постоянно и потому даже за короткий промежуток времени информация о сканировании безнадежно устаревает, следовательно, необходимо вместе с контрольной суммой сохранять и дату последнего времени сканирования (отображая ее пользователю), а так же предусмотреть кнопочку "rescan";
- многие "честные" файлы (особенно входящие в состав операционной системы) снабжены цифровой подписью или их целостность может быть проверена путем обращения к серверам Microsoft, что осуществляется _намного_ быстрее антивирусного сканирования и, если по данным Microsoft, файл не изменен, зачем его прогонять через антивирусы?!
- алгоритм CRC32 походу может показаться плохой идей, поскольку выдает множество коллизий (т. е. разные файлы имеют идентичные контрольные суммы) к тому же его легко

умышленно подделать, модифицировав _любое_ количество байт файла и затем скорректировав 4е байта так, чтобы "скомпенсировать" искажения. однако! CRC32 очень шустро работает, обгоняя MD5 и другие "хорошие" алгоритмы, поэтому возникает следующая идея: для каждого файла, прогоняемого через антивирусы, мы генерируем CRC32 и MD5 (с учетом времени сканирования, накладными расходами на расчет контрольной суммы можно пренебречь), а вот при последующей проверке залитого пользователем файла сначала проверяем CRC32 (а проверяется он очень быстро) и если такой контрольной суммы в нашей базе нет, то MD5 можно и не вычислять — зачем? ведь и так видно, что данный файл еще не проверялся;

Так же крайне желательно реализовать опцию, позволяющую пользователю выбирать режим сканирования с эвристикой и без (чего не сделано на virus-total). Эвристика представляет собой довольно затратную по времени и ресурсам ЦП операцию, но далеко не все пользователи доверяют полученным результатам и хотят видеть имя конкретного вируса (если он есть), а не расплывчатое предупреждение, обычно ругающееся на упаковщик/протектор, которым обработан честный файл. С другой стороны, вирусописателям совершенно неинтересно сканирование по базе (т. к. только что написанного вируса там _заведомо_ нет) и они предпочли бы задействовать только эвристику, экономя тем самым ресурсы _нашего_ сервера. Так почему бы не пойти им навстречу?!

Закачка больших файлов предоставляет серьезную проблему, имеющую несколько решений: самое простое (и самое глупое) установить верхний предел закачиваемого файла в пару мегабайт (или около того), чуть-чуть умнее: лимитировать суммарный размер всех файлов, закаченных за сутки с данного IP (но тут возникает проблема определения этого самого IP, поскольку очень часто мы будет видеть не IP пользователя, а IP прокси сервера провайдера). Полезно порекомендовать пользователем сжимать файлы перед отправкой zip'ом или другим популярным архиватором для уменьшения нагрузки на канал или делать это автоматически на клиентской стороне специальным скриптом. Наконец, за сканирование больших файлов можно взимать деньги, но об этом мы поговорим чуть позже, а пока продолжим тему оптимизации.

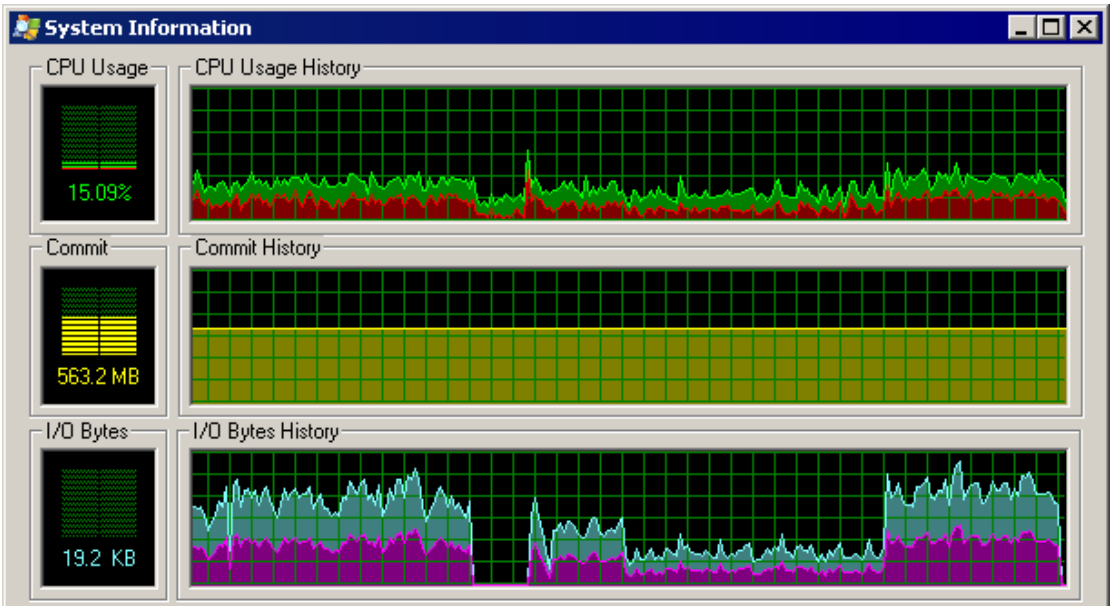

Рисунок 3 при онлайновом антивирусном сканировании основная нагрузка ложится на подсистему ввода/вывода и оперативную память. требования к мощности процессора не столь значительны (антивирусы располагались на жестком диске и работали в режиме чистого сканера без эвристического анализатора)

Профилировка показывает, что львиная доля накладных расходов приходится на запуск антивируса, инициализацию его движка и загрузку антивирусных баз. Перемещение антивирусов на виртуальный диск _существенно_ увеличивает "подвижность" системы, но накладные расходы на создание новых процессов по прежнему будут большим, поэтому, мы либо используем GUI-версии антивирусов и, путем эмуляции клавиатурного ввода, воздействует на элементы управления, заставляя их сканировать новые файлы и выдавать

результат. При этом антивирус запускается всего один раз! Красота! Впрочем, можно реализовать и динамический алгоритм: при небольшой нагрузке на сервер о накладных расходах на порождение новых процессов можно не заботиться, а с ростом нагрузки — просто брать несколько файлов, закаченных пользователи за последние несколько минут и "скармливать" их антивирусу одним скопом, в результате чего количество запусков последнего резко сокращается. Главное не запутаться какой пользователь что закачал, но это уже мелочи технической реализации.

Естественно, запускать сканирование лучше на всех антивирусах параллельно, а не последовательно и вместо того, чтобы "тупо" запрещать пользователю закрывать окно браузера до окончания процесса сканирования (как это делает virus-total), отслеживать TCP/IP соединение и при его обрыве, автоматически "выбрасывать" файл, принадлежащий данному пользователю из очереди файлов, поставленных на сканирование. Плюс реализовать стандартную кнопку "отмены" (так же отсутствующую у virus-total'a), потому как если пользователь видит, что первые три-четыре антивируса ничего не находят, так следует ли дожидаться результатов проверки _всех_ антивирусов?! Особенно, если самые качественные антивирусы поставить вперед остальных, выделив им максимальный приоритет ЦП (как вариант, можно вообще не следить за TCP/IP сессий и при заливке нового файла назначать пользователю ID задачи, который он может ввести в любое время, отключившись от сети и повторно подключившись, например, через час, когда его очередь уже подошла, так же можно рассылать результаты сканирования по email, — все это позволяет пользователю не скучать в ожидании приближения его очереди).

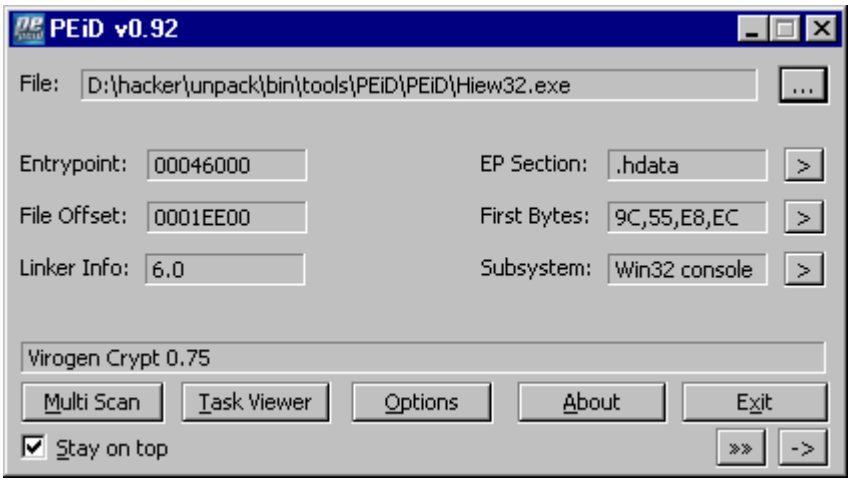

Рисунок 4 внешний вид утилиты PEiD, определяющий тип и версию упаковщика/протектора исполняемых файлов

И совсем не помешать прикрутить к нашему сервису утилиту вроде PEiD, определяющую тип и версию упаковщика/проектора (правда, довольно часто ошибающуюся) и (опционально) реализовать распаковку набором статических распаковщиков, работающих _намного_ быстрее распаковщиков, встроенных в антивирусы, однако, тут есть один подводный камень — хотя 99% вирусов распознаются по распакованному дампу, некоторые, особо ленивые сотрудники антивирусных компаний, включают в базу сигнатуры _упакованного_ файла и после распаковки он перестает опознаваться как вирус, однако, учитывая, что распакованный файл прогоняется через легион антивирусов, вероятность ложно-негативного срабатывания стремиться к нулю.

юридические проблемы лицензирования

Пользовательские соглашения (EULA) на коммерческие антивирусы не разрешают использовать их в онлайновых сервисах без заключения специальных контрактов, что вообще-то логично. Однако, не стоит думать, что всякий контракт _обязательно_ связан с необходимостью выплаты дополнительных отчислений. Вовсе нет! Достаточно проявить надлежащий дипломатический подход!

Крупные брэнды заинтересованы в рекламе своей продукции и потому охотно разрешают использовать полнофункциональные версии антивирусов без всяких отчислений, поскольку, конечный пользователь реально видит кто сосет, а кто нет. Правда, они могут

выдвинуть встречные условия типа сохранения логотипов, генерации ссылок на их сайты и т. д., однако, все это уже мелочи решаемые в рабочем порядке.

Мелкие брэнды… они, конечно, понимают, что сравнение с конкурентами будет не в их пользу, однако, тут есть один очень интересный момент. Мелкие антивирусные компании страдают хронической нехваткой свежих вирусов, которые попадают к ним в последнюю очередь и потому онлайновый сервис, автоматически отсылающий им вирусы, уже детектирумые конкурентами — прекрасное средство пополнения своих вирусных баз и продвижения в различных рейтингах. С мелких брэндов даже можно взимать плату за каждого нового неизвестного им вируса и они в своей массе согласны платить!

Короче говоря, лицензионные проблемы — это и не проблемы вовсе. А вот проволочек тут предостаточно и нужно заранее быть готовыми к тому, что нас будут перебрасывать от одного ответственного лица (которое ни хвоста не решает) к другому, третьему, и так по цепочке… Но это уже издержки цивилизации, против которых не попрешь.

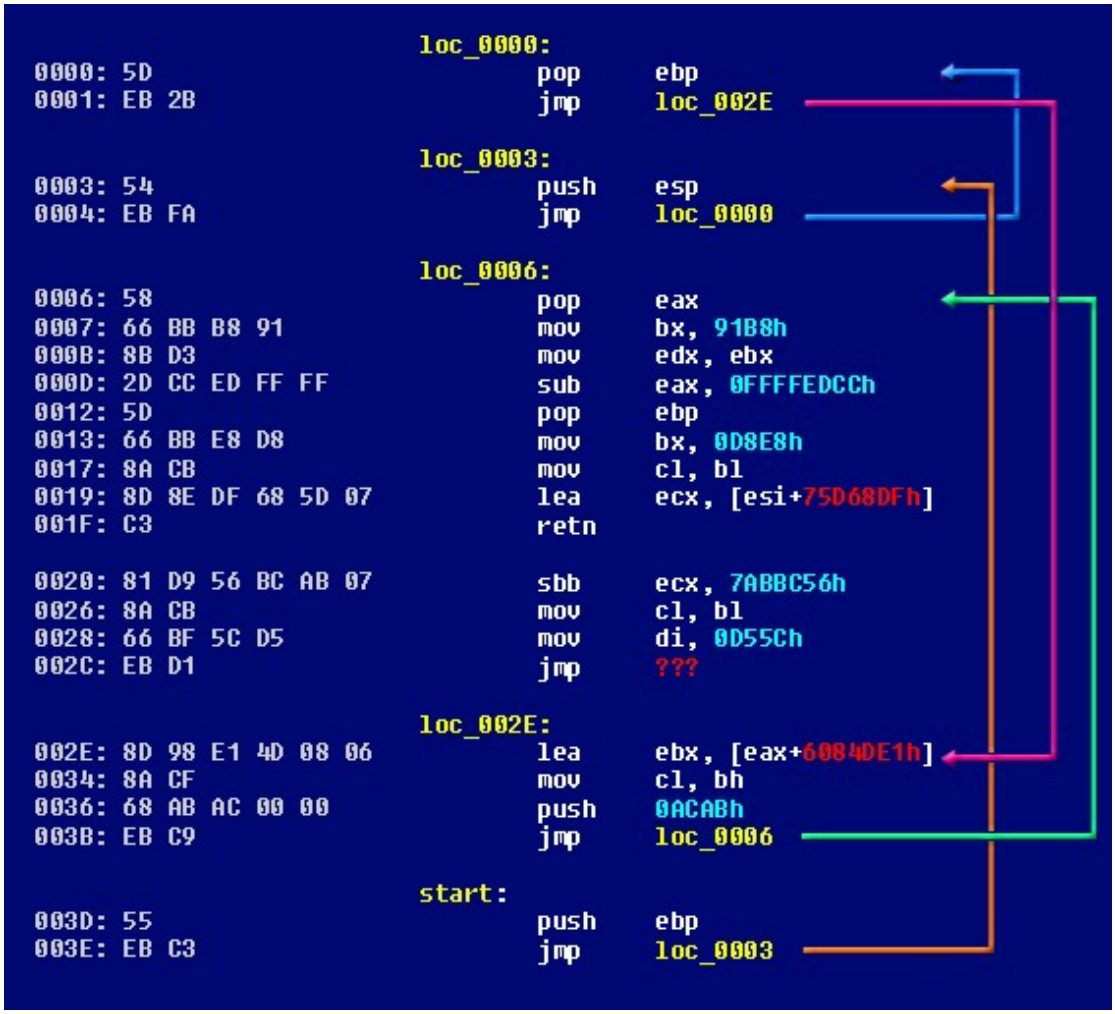

Рисунок 5 внутри дизассемблерного кода вирусного тела

>>> врезка зарабатываем деньги лопатами

Мир жесток и все в этом мире упирается в деньги. На голом энтузиазме никой онлайновый сервис долго не продержится, поэтому, приходится разрабатывать не только программный код, но и жизнеспособную бизнес-схему.

Рассмотрим возможные источники дохода. Первое — рост посещаемости нашего сайта, а на посещаемости, как известно, можно нехило заработать, особенно, если мы, например, продаем собственные защитные комплексы, предлагаем услуги по пен-тестингу, etc. Онлайновый сервис привлекает клиентов намного активнее любых баннеров и, что самое главное, он привлекает именно тот контингент потребителей, который нам нужен,

следовательно, возросшие объемы продаж покроют все расходы на поддержку и обслуживание серверов, отплату трафика и т. д.

Второе — отчисления от антивирусных компаний, чью продукцию мы рекламируем и кому передаем штаммы свежих вирусов. Тут, правда, особо много не заработаешь, поскольку, лишь небольшая часть посетителей нашего сайта кликнет по ссылке "купить" антивирус, а стоимость одного вирусного штамма обычно составляет $1 или менее того. Вот и считайте, на какой уровень посещаемости нужно выйти, чтобы окупить расходы на поддержку сервера, которые, кстати говоря, тем больше, чем выше посещаемость.

Третье — взимать пиастры непосредственно с самих пользователей. Хочешь подолгу стоять в очередях и сканировать файлы не больше чем …. мегабайт, пожалуйста — пользуйся нашим сервисом бесплатно! Хочешь иметь определенные привилегии — будь добр заплатить. Главное — выбрать удобную схему оплаты. Здесь вам не Америка, здесь климат (финансовый) иной. Кредитные карты имеют единицы, электронные системы платежей только начинают набирать популярность. Зато практически каждый IT-специалист имеет сотовый телефон, а это значит, что можно либо воспользоваться микро платежами через SMS, либо потребовать от клиента сообщить номер карты универсальной оплаты, перечислив заданную сумму на его счет, который он может расходовать, пользуясь нашим сервисом. Как показывает практика, сотовые платежи приносят наибольшую отдачу, поскольку, телефоны распространены повсеместно, а сам процесс оплаты требует минимум телодвижений и (что тоже немаловажно) клиент практически ничем не рискует. Ну, допустим, передаст он номер карты на 150 рэ, а мы пошлем его в пешее эротическое путешествие — он же фактически ничего (кроме настроения) не теряет. А вот с кредитными картами все намного сложнее и есть риск, что нечестный оператор снимет с них совсем не ту сумму, которая ожидалась. Тоже относится и к микроплатежам через SMS. Гарантий, что снимут именно 150 рэ, а не 450 рэ — у клиента нет никаких, а раз так…

>>> врезка проблемы конфиденциальности

Вирусы встречаются не только в программах, но так же офисных документах, pdf'ах и прочих файлах с конфиденциальной информацией, разглашать которую крайне нежелательно, поэтому, необходимо предусмотреть опцию "не отсылать данный файл в антивирусные центры", при необходимости взводимую пользователем. Технически это реализуется проще простого, но… как избежать злоупотреблений?! Особенно если мы строим наш бизнес на отправке свежих штаммов разработчикам антивирусов?

Идея первая (тупая до безобразия) — наплевать на все приличия, и отсылать файлы в антивирусные центры независимо от состояния каких-то там галочек. Главное — создать у пользователя иллюзию, что его конфиденциальность строго блюдут, ну а что происходит _на_ _самом_ _деле_ он все равно не узнает. Ну… до тех пор, пока тайное не станет явным и не разразиться скандал, идущий совсем не на пользу нашему ресурсу.

Идея вторая — поддерживать эту опцию только для пользователей, открывшим у нас счет, что, кстати говоря, представляет собой нехилую мотивацию для оплаты услуг, особенно с учетом того, что о конфиденциальности главным образом беспокоятся корпоративные пользователи, привыкшие платить за услуги, в отличии от домашних юзеров, тяготеющих к халяве все зависимости от стоимости полнофункционального аккаунта.

Идея третья — файл все-таки отправлять, но перед этим удалять всю текстовую и графическую информацию, что, кстати говоря, не противоречит ни логике, ни здравому смыслу, ни даже соглашению, заключенному с пользователем нашего сервиса.

заключение

Разумеется, в данной статье охвачены далеко не все проблемы, с которыми неизбежно столкнется всякий, попытавшийся воздвигнуть подобный онлайновый сервис, однако, мыщъх предоставил вполне законченную, отлаженную и работоспособную схему, которая скоро будет запущена в промышленную эксплуатацию.

Сразу же хотелось бы уточнить два момента: а) никаких корпоративных тайн мыщъх не раскрыл и раскрывать не собирается; б) запросы на консультацию и предложения о реализации аналогичных движков идут лесом, поскольку, у меня нет ни малейшего желания дважды решать одну и туже задачу. Тем не менее, мыщъх всегда открыт для общения и сотрудничества (контактный адрес заинтересованные лица могут добыть в редакции).

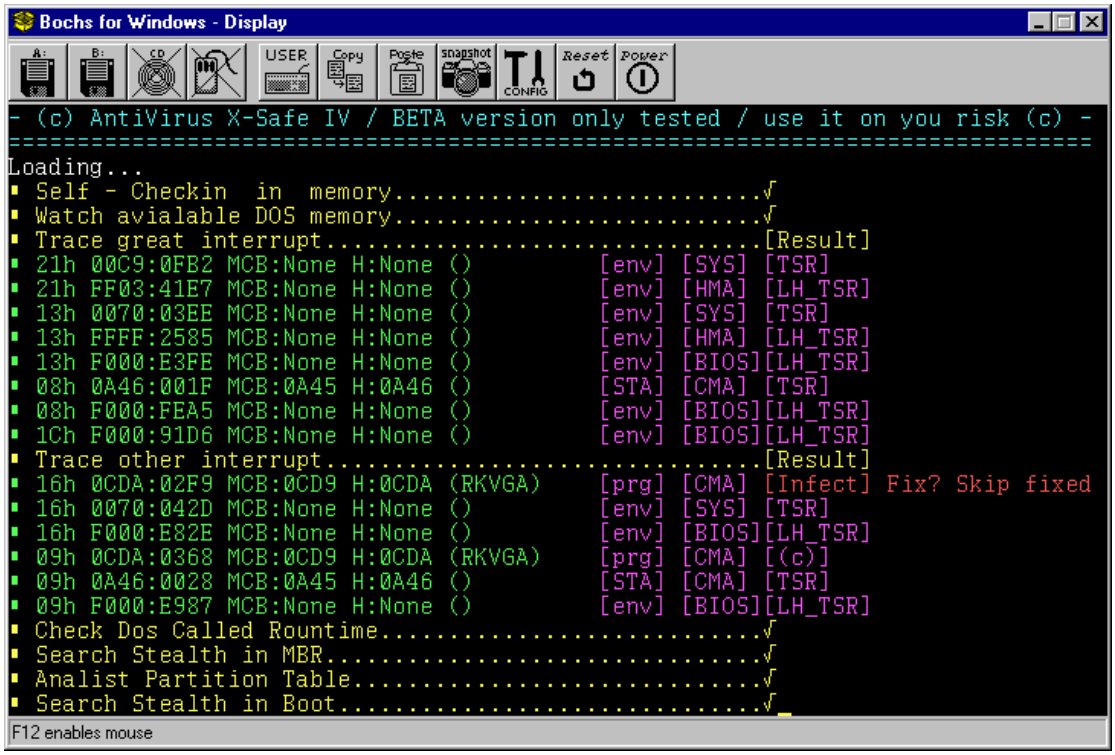

Рисунок 6 XSafe – один из первых антивирусов, разработанных мыщъх'ем еще под MS-DOS, в настоящее время представляющей не более чем исторический интерес и требующей для своей работы эмулятора ;)