у google под колпаком
tagline: интернет как альтернативна гестапо

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

право на privacy (защиту личной информации) уже и без того изрядно потрепанное в боях подверглось очередной атаке. на этот раз со стороны гиганта Google, шпионящего за нами с помощью закладок, встроенных в популярные браузеры (Горящий Лис, Опера), а так же панель Google Toolbar, установленную у миллионов пользователей. как обнаружить факт шпионажа (мыщъх, например, обнаружил это чисто случайно), какая именно информация передается, чем это нам грозит в практическом плане и можно ли предотвратить разгул безобразия своими собственными силами?

введение

Интернет представляет собой отличный инструмент для контроля за деятельностью его обитателей, в котором заинтересованы и правительственные учреждения, и крупные/мелкие корпорации, ну и не в последнюю очередь хакеры. Стандартный браузер и так предоставляет слишком много информации о клиенте, передавая ее узлу с которым осуществляется соединение: тип и версия операционной системы, тип и версия самого браузера, а так же адрес предыдущей посещенной страницы. Невероятная богатая информация для статистического анализа, но, к сожалению аналитиков (и к счастью простых пользователей), полностью децентрализованная и разобщенная: не существует никакого единого центра по сбору данных и хотя некоторые фирмы предоставляют бесплатные счетчики (типа www.SpyLog.ru), они контролируют лишь те сетевые ресурсы, на которых они установлены.

Панель управления Google-Toolbar (см. рис 1), выпущенная для Горящего Лиса и IE, не только упрощает web-серфинг, но и передает Googl'у информацию о посещаемых узлах, типе и версии браузера/операционной системы, честно предупреждая об этом в пользовательском соглашении, поэтому тут никаких претензий у нас нет. По официальной версии, полученные данные не разглашаются, не передаются никаким третьим лицам (типа ФБР), а используется исключительно для улучшения качества поиска. Наиболее часто посещаемые ссылки получают более высокий приоритет и выводятся первыми, от чего выигрывает как сама поисковая машина, так и конечные пользователи (независимо от того, установлен ли у них Google Toolbar или нет).

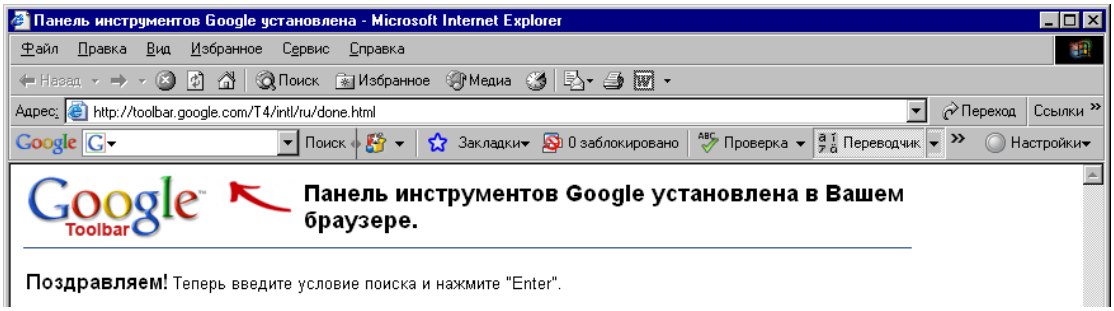

Рисунок 1 панель инструментов Google Toolbar – легальный шпионский компонент

Естественно, далеко не каждый готов делиться с Google какой бы то ни было личной информацией, поэтому, Google вступила в сговор с разработчиками некоторых популярных браузеров (Горящий Лис, Опера), убедив их встроить специальные закладки, скрытно передающих информацию о всех действиях пользователя в специальных аналитический центр, принадлежащий Googl'у. Быть может, в этом и кроется секрет "бесплатности" Оперы? Как знать….

Ладно, не будем гадать на кофейной гуще, а лучше пронаблюдаем за процессом передачи данных своими собственными глазами и подумаем как предотвратить утечку персональной информации, раскрытие которой может иметь далеко идущий характер. В частности, хакеры давно и небезуспешно используют Google для атак на сайты, о чем можно

прочесть в статье: "Google Hacking for Penetration Testers", написанной хакером по прозвищу Johnny Long: www.blackhat.com/presentations/bh-europe-05/BH_EU_05-Long.pdf

Больше всего, конечно, от этого страдают владельцы web-серверов, но и обычным пользователям временами приходится несладко. Так что… подтолкнем под себя хвост и займемся экспериментами, подробно описанных мыщъх'ем шаг за шагом.

что нам понадобиться

Для проведения экспериментов нам понадобиться: **Горящий Лис** (версия 1.5), **Internet Explorer** (версия 6.0.2800.1106), **Опера** (версия 8.51). Остальные версии мыщъх не проверял, поэтому их поведение может отличаться от описанного.

Еще нам понадобиться **sniffer** (грабитель сетевого трафика) и **брандмауэр** (для защиты от утечек информации). Мыщъх использует **SyGate Personal Firewall** компании SyGate (ныне скупленной корпорацией Symantec), включающий в себя неплохой пакетный фильтр. До версии 4.2 для некоммерческого использования он был бесплатен, то теперь за полную версию просят денежку, а из демонстрационной пакетный фильтр исключен, поэтому, приходится либо раскошеливаться, либо искать антиквариат, либо использовать какой-нибудь другой брандмауэр плюс бесплатный **tcpdump** (http://www.tcpdump.org/), портированный под множество операционных систем, среди которых значится и Windows.

Так же, для чистоты эксперимента рекомендуется установить свой собственный web-сервер, чтобы исключить все побочные воздействия. Мыщъх использует **Small Http Server** (http://smallsrv.com/), который советует и остальным, тем более, что для граждан бывшего СНГ он бесплатен.

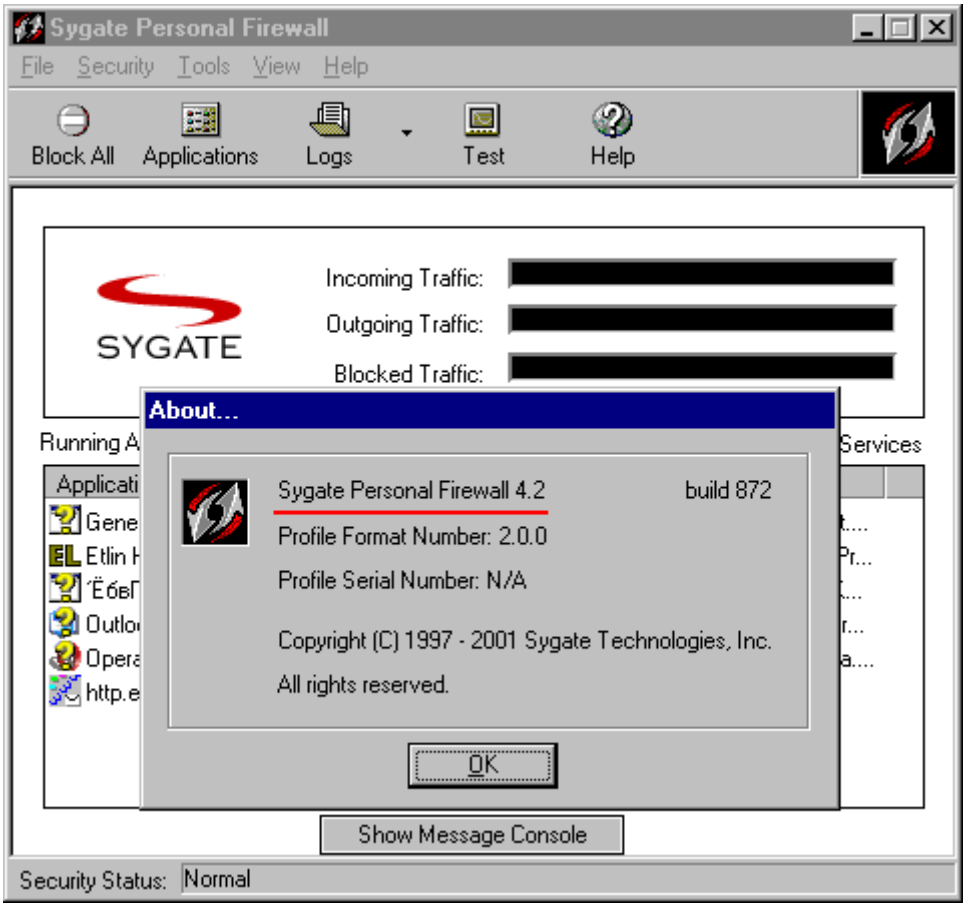

Рисунок 2 внешний вид персонального брандмауэра

взятие Горящего Лиса с поличным

Берем свежеустановленного Горящего Лиса за хвост и ходим по адресу http://nezumi.org.ru (адрес мыщъх'иного web-сервера), открываем SyGate Personal Firewall, лезем

в Logs → Packet Log (при этом галочка "Capture Full Packet" в "Options → Log" должна быть заблаговременно установлена) и видим, что в пакетном логе появились какие-то странные и совершенно левые IP-адреса (см. рис. 3), с которыми сношался процесс firefox.exe через 80'й порт (локальный адрес мыщъх'ного сервера в логе отсутствует, поскольку пакетный фильтр Sygate Personal Firewall'a игнорирует трафик идущий через loop-back петлю 127.0.0.1).

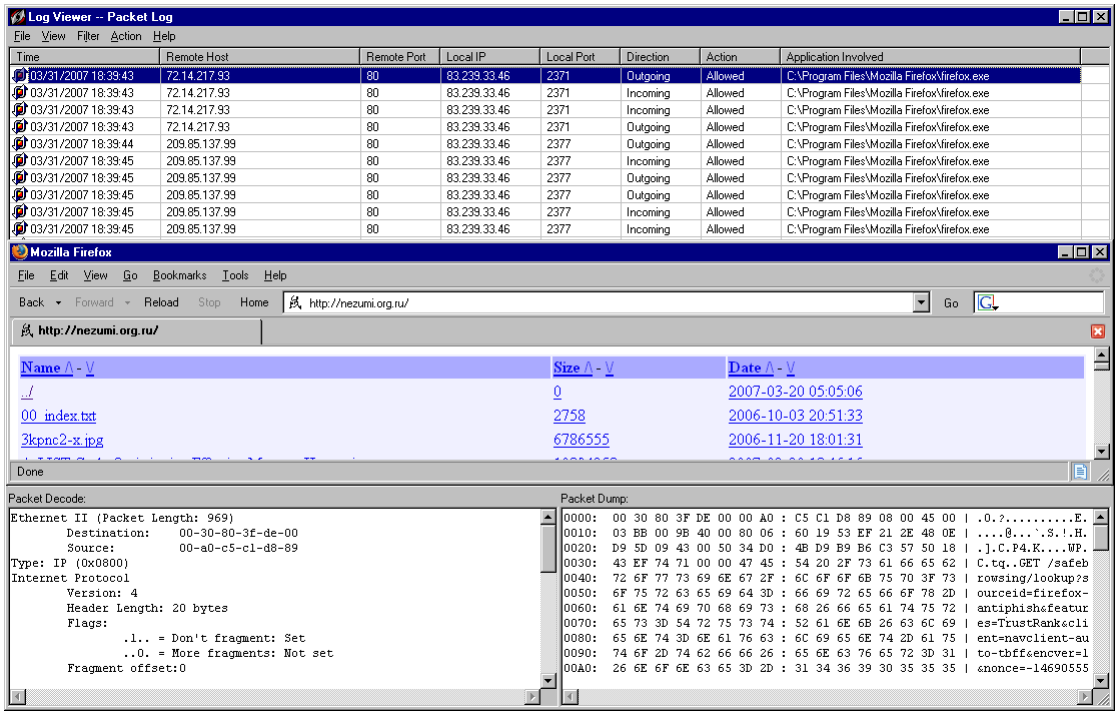

Рисунок 3 при заходе на сервер http://nezumi.org.ru Горящий Лис тайком передает какую-то информацию по адресам 72.14.217.93 и 209.85.137.99, принадлежащий корпорации Google

Попробуем выяснить кому принадлежат эти IP, определив их доменное имя посредством штатной утилиты tracert.exe, стараниями которой мы быстро узнаем, что адресу 72.14.217.93 соответствует доменное имя **bu-in-f93.google.com** (см. рис. 4), а 209.85.137.99 – **mg-in-f99.google.com** (см. рис. 5).

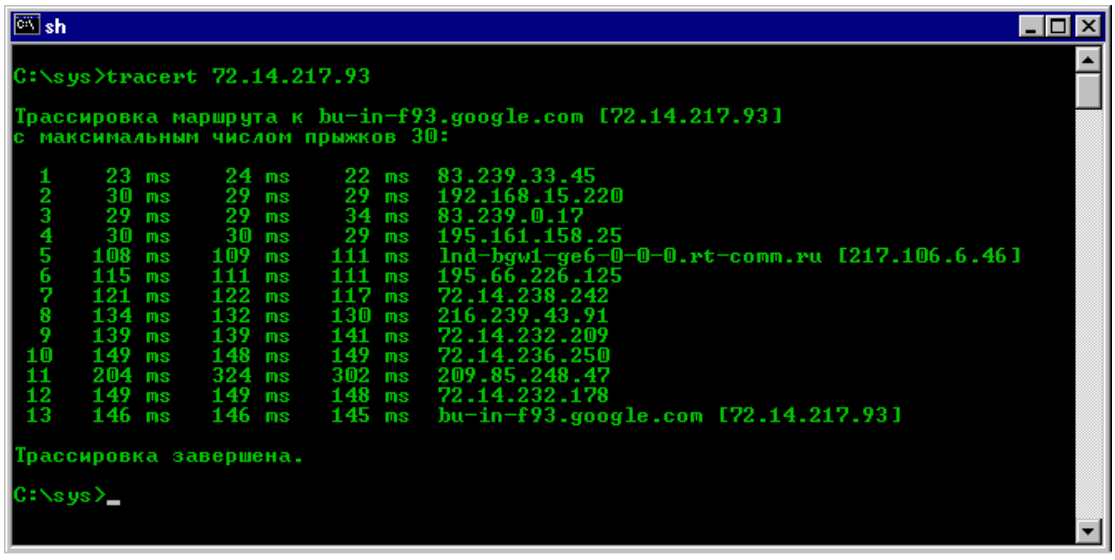

Рисунок 4 IP-адресу 72.14.217.93 соответствует доменное имя bu-in-f93.google.com

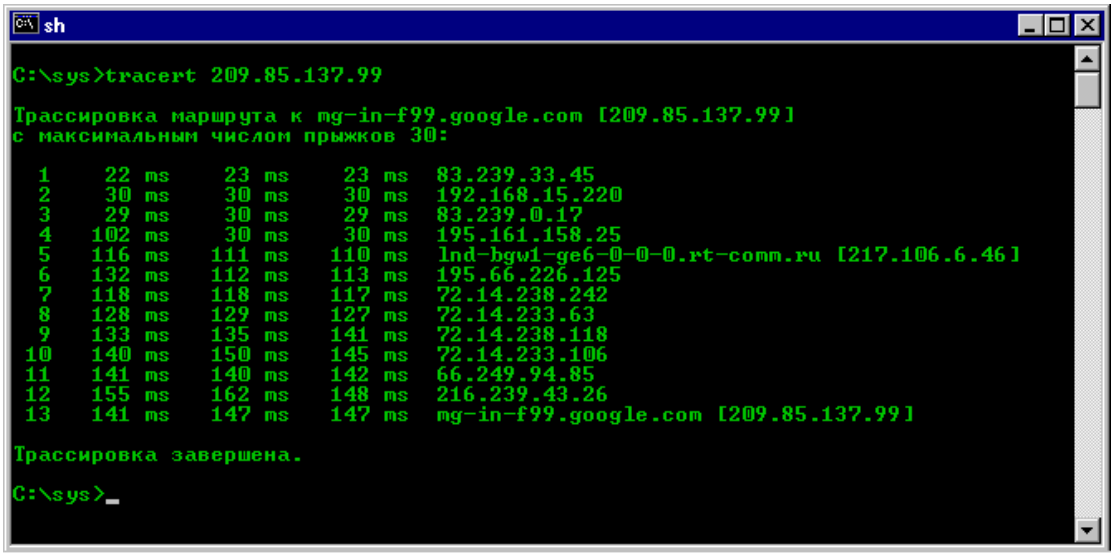

Рисунок 5 IP-адресу 209.85.137.99 соответствует доменное имя mg-in-f99.google.com

Ага! В воздухе уже запахло паленным (хвостом Горящего Лиса). Оба адреса принадлежат корпорации Google и, что самое интересное, находятся в различных подсетях. Короче, факт скрытой передачи персональных данных можно считать надежно установленным. Остается только выяснять, _что_ именно за информация передается. Это легко! Достаточно взглянуть на окно дампа, содержимое которого приведено ниже:

```
0000:  00 30 80 3F DE 00 00 A0 : C5 C1 D8 89 08 00 45 00 | .0.?..........E.
0010:  03 FF F8 41 40 00 80 06 : 68 30 53 EF 21 2E 48 0E | ...A@...h0S.!.H.
0020:  D9 5B 08 95 00 50 AC 7C : 5B 7B 32 F6 FD E4 50 18 | .[...P.|[{2...P.
0030:  41 6A 05 C4 00 00 47 45 : 54 20 2F 73 61 66 65 62 | Aj....GET /safeb
0040:  72 6F 77 73 69 6E 67 2F : 6C 6F 6F 6B 75 70 3F 73 | rowsing/lookup?s
0050:  6F 75 72 63 65 69 64 3D : 66 69 72 65 66 6F 78 2D | ourceid=firefox-
0060:  61 6E 74 69 70 68 69 73 : 68 26 66 65 61 74 75 72 | antiphish&featur
0070:  65 73 3D 54 72 75 73 74 : 52 61 6E 6B 26 63 6C 69 | es=TrustRank&cli
0080:  65 6E 74 3D 6E 61 76 63 : 6C 69 65 6E 74 2D 61 75 | ent=navclient-au
0090:  74 6F 2D 74 62 66 66 26 : 65 6E 63 76 65 72 3D 31 | to-tbff&encver=1
00A0:  26 6E 6F 6E 63 65 3D 2D : 31 34 37 30 36 37 37 35 | &nonce=-14706775
```

Листинг 1 информация, передаваемая Лисом узлу bu-in-f93.google.com

Даже неспециалисту понятно, что bu-in-f93.google.com представляет собой web-сервер, которому Лис посылает запрос "`GET /safebrowsing/lookup?sourceid=firefox-antiphish&features=TrustRank&client=navclient-auto-tbff&encver=1`", вызывающий скрипт `/safebrowsing/lookup` и передающий ему строку параметров, включающую в себя среди прочей интересной информации зашифрованный URL, посещаемой страницы вместе с типом/версией браузера/операционной системы, передаваемых открытым текстом.

А вот что передается узлу mg-in-f99.google.com:

```
0000:  00 30 80 3F DE 00 00 A0 : C5 C1 D8 89 08 00 45 00 | .0.?..........E.
0010:  03 A9 01 77 40 00 80 06 : 26 02 53 EF 21 2E D1 55 | ...w@...&.S.!..U
0020:  89 63 09 49 00 50 F7 E4 : 17 E3 89 73 B5 75 50 18 | .c.I.P.....s.uP.
0030:  44 70 3E 02 00 00 47 45 : 54 20 2F 73 65 61 72 63 | Dp>...GET /searc
0040:  68 3F 73 6F 75 72 63 65 : 69 64 3D 6E 61 76 63 6C | h?sourceid=navcl
0050:  69 65 6E 74 2D 66 66 26 : 66 65 61 74 75 72 65 73 | ient-ff&features
0060:  3D 52 61 6E 6B 26 63 6C : 69 65 6E 74 3D 6E 61 76 | =Rank&client=nav
0070:  63 6C 69 65 6E 74 2D 61 : 75 74 6F 2D 66 66 26 67 | client-auto-ff&g
0080:  6F 6F 67 6C 65 69 70 3D : 4F 3B 32 30 39 2E 38 35 | oogleip=O;209.85
0090:  2E 31 33 37 2E 39 39 3B : 32 33 30 26 63 68 3D 38 | .137.99;230&ch=8
00A0:  35 31 37 38 31 32 64 37 : 26 71 3D 69 6E 66 6F 3A | 517812d7&q=info:
00B0:  68 74 74 70 25 33 41 25 : 32 46 25 32 46 6E 65 7A | http%3A%2F%2Fnez
00C0:  75 6D 69 2E 6F 72 67 2E : 72 75 25 32 46 20 48 54 | umi.org.ru%2F HT
00D0:  54 50 2F 31 2E 31 0D 0A : 48 6F 73 74 3A 20 74 6F | TP/1.1..Host: to
```

Листинг 2 информация, передаваемая Лисом узлу mg-in-f99.google.com

Отчетливо виден ничем не прикрытый запрос "`GET /search?client=navclient-auto&googleip=O;1532&iqrn=TPVB&orig=0n2ln&ie=UTF8&oe=UTF-8&querytime=Q0B&features=Rank:&q=info:`**`http%3a%2f%2fnezumi%2eorg%2eru`**`&ch=702785874955 HTTP/1.1`", с незашифрованным адресом http%3a%2f%2fnezumi%2eorg%2eru, что в переводе на нормальный язык выглядит как: http://nezumi.org.ru.

Факт утечки информации налицо!!! А теперь (разнообразия ради) попробуем установить панель Google toolbar и посмотрим как она повлияет на конечный результат. Идем на http://www.google.com/intl/en/options/, находим "Toolbar — Add a search box to your browser", и устанавливаем версию, разработанную специально для Горящего Лиса http://www.google.com/tools/firefox/index.html, убедившись, что галочки напротив пунктов "PageRank Display" и "Safe Browsing" (см. рис. 6) находятся во взведенном состоянии (tools → extensions → google toolbar for fire fox → options), в противном случае персональная информация никуда передаваться не будет! (активность Google toolbar _никак_ _не_ зависит от того, отображается ли она на панели инструментов или нет).

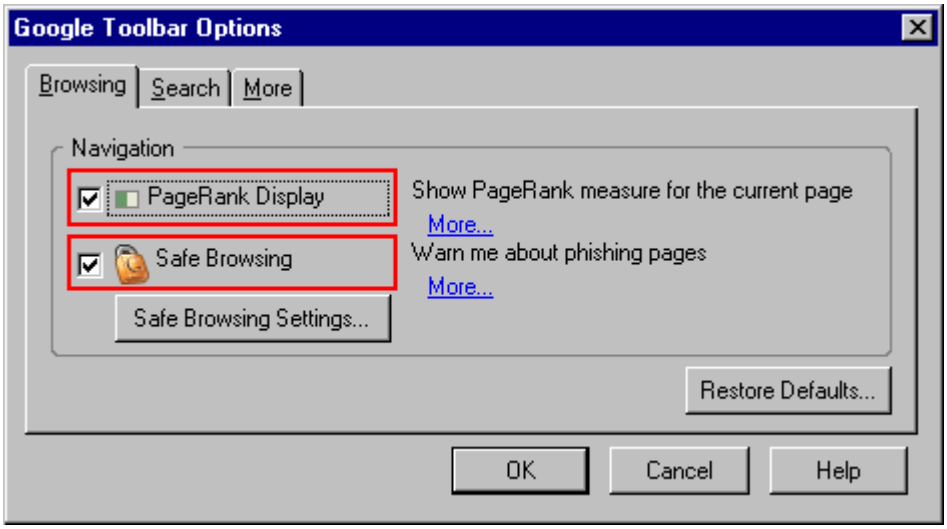

Рисунок 6 настройки Google Toolbar, ответственные за передачу персональной информации

ОК, повторяем попытку захода на http://nezumi.org.ru и смотрим в пакетный лог (см. рис. 7), в котором теперь вместо бессловесных IP-адресов появляются доменные имена **sb.google.com** и **www.google-analytics.com**, первое из которых соответствует bu-in-f93.google.com, а второе — mg-in-f99.google.com, что легко определить, изучив протокол обмена и сравнив его с предыдущим результатом.

Другими словами, в Горящего Лиса _изначально_ встроено ядро панели Google toolbar, причем без возможности его отключения штатными средствами (вариант с правкой исходных текстов не предлагать).

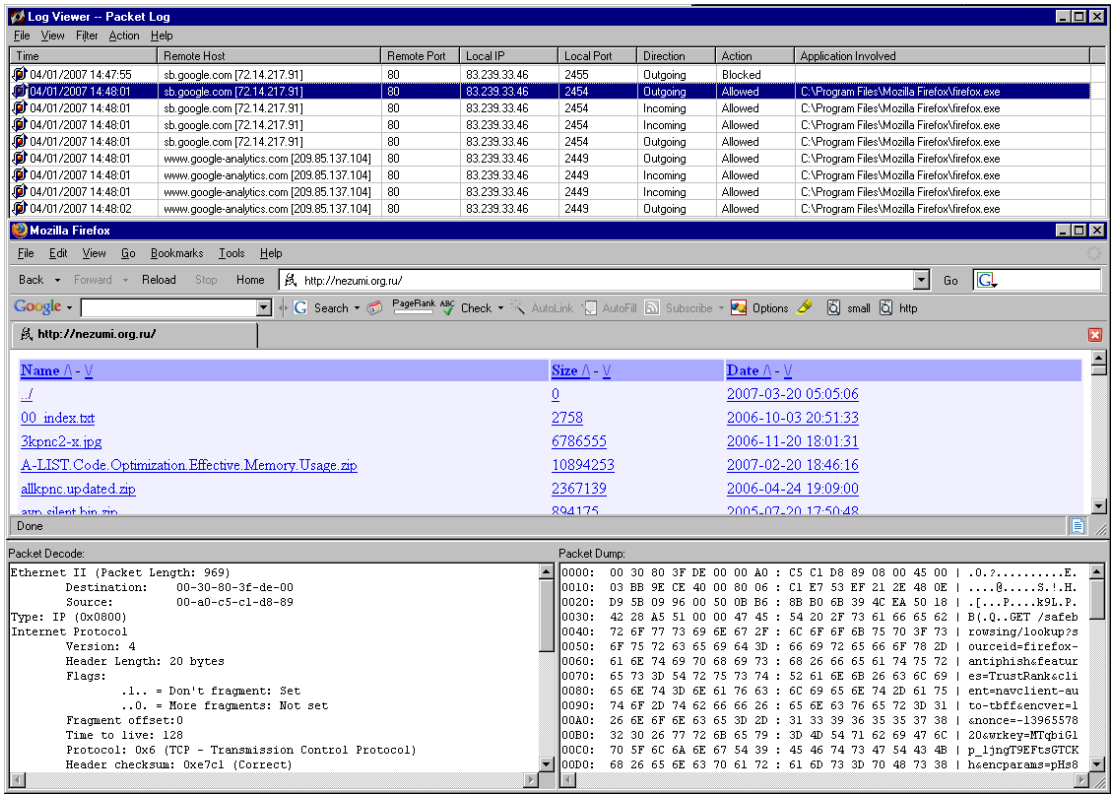

Рисунок 7 Google toolbar передает персональную информацию узлам sb.google.com и www.google-analytics.com, на которых установлены те же самые скрипты /safebrowsing/lookup и /search, что и на f93.google.com, mg-in-f99.google.com

А теперь смертельный номер! Заходим в настройки Google toolbar и _отключаем_ "PageRank Display" и "Safe Browsing", после чего передача персональной информации тут же _прекращается_ и это хорошо! То есть, ***чтобы предотвратить утечку персональной информации, необходимо установить Google toolbar, залезть в настройки и отключить "PageRank Display" и "Safe Browsing"***

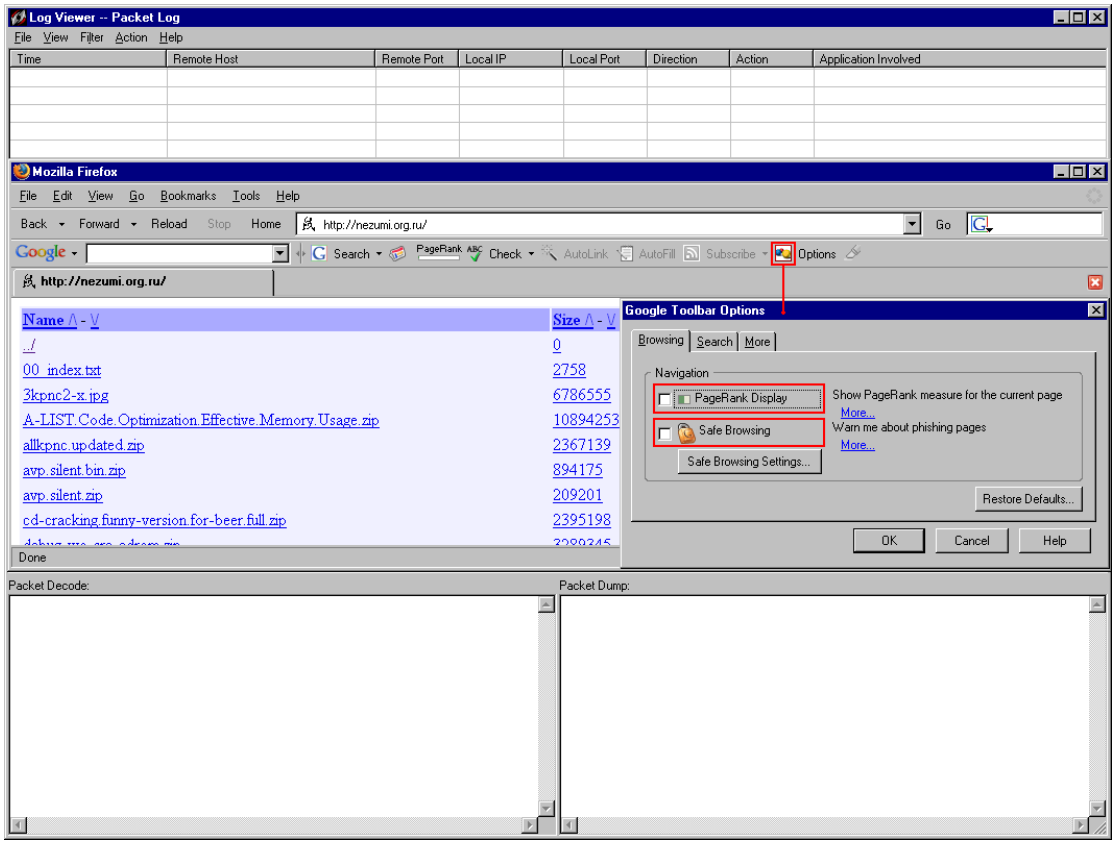

Рисунок 8 персональная информация прекращает утекать при отключении PageRank Display и Safe Browsing

Аналогичная закладка имеется и в Опере, но в силу отсутствия для нее специальной версии Google toolbar, утечку персональной информации предотвратить не так-то просто (мыщхъ это сделал путем бит-хака, то есть хирургического вмешательства в двоичный код, однако существуют и другие методы, которые мы обсудим чуть позже).

В Internet Explorer'е закладки от Googl'а отсутствуют (еще бы! ведь Google и Microsoft заклятые враги-конкуренты), однако, поскольку Internet Explorer — это сплошная дыра (типа "друшлак"), то по соображениям безопасности пользоваться им категорически не рекомендуется.

кто стучит на тебя?

Берем пропатченную Оперу, Internet Explorer или любой другой браузер _заведомо_ не содержащий закладок и совершаем марш-бросок на http://subscene.com/ где нажимаем ссылку "Search" (см. рис. 9) и смотрим в пакетный лог (см. рис. 10).

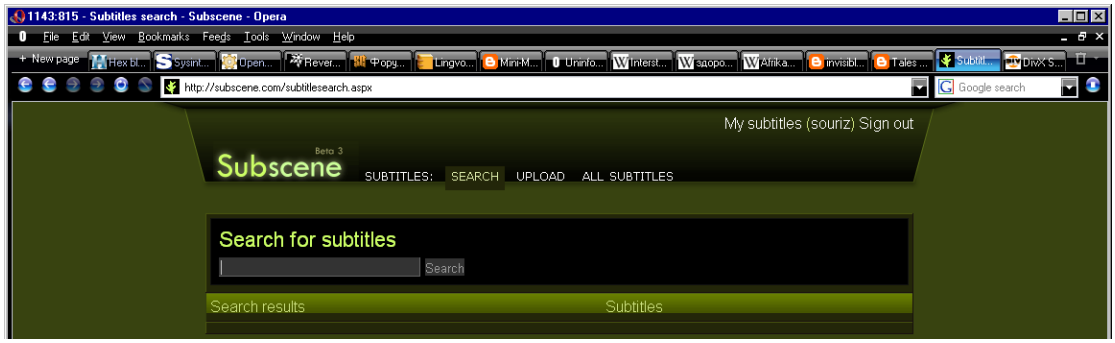

Рисунок 9 просмотр сайта http://subscene.com/subtitlesearch.aspx посредством браузера, _заведомо_ не "работающего" на Google

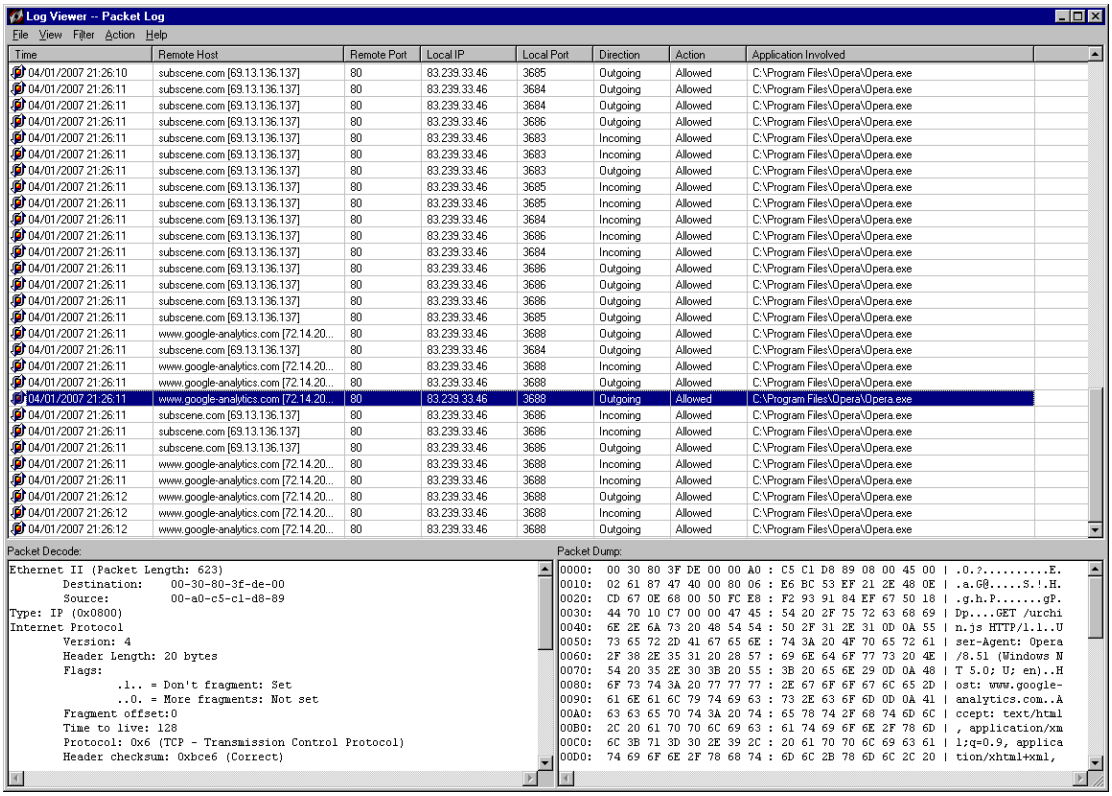

Рисунок 10 лог брандмауэра содержит большое количество обращений к узлу www.google-analytics.com

Что за черт?! Лог брандмауэра буквально кишит обращениями к узлу **www.google-analytics.com**, отсылая ему запросы "`GET /urchin.js HTTP/1.1`". Какая это су… сумчатая кенгуру стучит на нас?! И откуда взялся urchin.js?! Это что-то новенькое! Раньше такого не встречалось!

Просмотр исходного кода HTML-страницы быстро выявляет следующий Java-Script, код которого и является стукачом:

```
<script src="http://www.google-analytics.com/urchin.js" type="text/javascript">
</script>
```

Листинг 3 Java-Script код в исходном HTML-коде, отправляющий персональную информацию в аналитический центр Google

Выходит, что subscene.com (как и многие другие web-узлы) активно сотрудничает с Google, добровольно передавая ему статистику нажатий на те или иные ссылки, а вместе с нею — персональную информацию о типе/версии браузера/операционной системе, языковых настройках и даже… локальном времени, что позволяет вычислить географическое местоположение посетителя.

чем это чревато?

Какой ущерб может нанести утечка персональной информации, стекающийся в аналитический центр корпорации Google при условии, что она не попадет к третьим лицам? Начнем с простых пользователей. Ну какое кому дело, кто куда ходит и на какие ссылки нажимает? Теоретически, Google может отслеживать посетителей "неправильных" сетевых ресурсов, пропагандирующих терроризм или распространяющих педофилию, но… ни одного подобного прецедента до сих пор зафиксировано не было! Тем не менее, это еще не означает, что можно и дальше бродить по сети и ничего не опасаться.

Рассмотрим типичную ситуацию — рядовую контору, сотрудники которой в "свободное от работы время" смотрят порнографию через https-proxy, шифрующие трафик, так что администратор даже и не догадывается какой гадостью занимаются его подопечные. Но, поскольку, передача персональной информации в аналитический центр Google осуществляется

в незашифрованном виде, то администратору достаточно всего лишь натравить greep на лог, чтобы все тайное немедленно стало явным.

Владельцам web-серверов приходится _намного_ хуже и утечка персональной информации приводит к _реальной_ угрозе нарушения безопасности. Создает, допустим, владелец web-ресурса виртуальную директорию, кладет в нее информацию-не-для-всех и дает ссылку заинтересованному лицу. Виртуальные директории не отображаются в списке содержимого каталога и чтобы добраться до них, нужно знать полный путь (фактически играющий роль пароля). Создавать виртуальные директории гораздо проще, чем заморачиватся с заведением новых пользователей и раздачей логинов/паролей (тем более, что далеко не всякий бесплатный хостер предоставляет подобную услугу, да и платный тоже). Кстати, аналогичного результата можно добиться, поместив файл в одну из "нормальных" директорий с запрещенным просмотром ее содержимого.

Но вся защита рушится, как только лицо, которому мы передали секретную ссылку, щелкнет по ней браузером, содержащим закладку или установленную панель Google toolbar, передающую URL в аналитический центр, направляющий содержимое виртуальной директории прямиком на индексацию, после чего _любой_ желающий может получить к ней доступ через поисковую машину Google!

Случай из личной жизни. Была у меня как-то сервере виртуальная папка /mp3/, доступная только пользователям с именем mp3 и таким же точно паролем, где лежала куча всякого добра, предназначенного сугубо для доступа в пределах домашней локальной сети (в самом деле, гонять файлы по витой паре намного удобнее, чем носиться с CD/DVD-RW дисками), и вот в один "прекрасный" день мыщъх заметил, что в приватную папку кто-то забрался и качает, причем не просто качает (как качал бы нормальный пользователь), а гребет все подряд, порядком напрягая канал. Глянул на адрес и обалдел — 82.208.10.16, в поле User-Agent которого без всякого стеснения и зазрения совести прямым текстом значилось: "Mozilla/5.0 (compatible; **Googlebot/2.1; +http://www.google.com/bot.html**)".

Каким же образом Google смог узнать логин/пароль к моей приватной папке?! Мысль о переборе мыщъх откинул сразу, а вот утечка информации через закладку, встроенную в Лиса, которым пользовался мыщъх, очень даже могла "настучать".

Таким образом, ***передавая ссылку на приватный ресурс лицу, пользующемуся Google toolbar (или браузером с закладкой внутри), мы рискнем, что назавтра об этом ресурсе узнает весь мир!!!***

>>> врезка сессии Google

Щелкая по ссылкам, выданным Googl'ом в ответ на наш запрос, мы передаем web-серверу информацию о текущей сессии, содержащую в поле referrer критерии запроса и соответствующий им результат. Вроде бы мелочь, а неприятно.

Сидит как-то мыщъх за монитором, жует бутерброд, и в ожидании пока IDA Pro дизассемблирует очередную программу, лениво поглядывает на консоль Small Http Server'a (см. рис. 11). Вдруг видит, как кто-то пытается утянуть phc3.full.pdf (электронная версия "записок мыщъх'а" целиком), причем, судя по строке referrer человек забрел явно с Googl'а, что весьма странно, поскольку мыщъх активно борется с Google, запрещая ему индексировать содержимое своего web-сервера по причинам, о которых мы говорили выше.

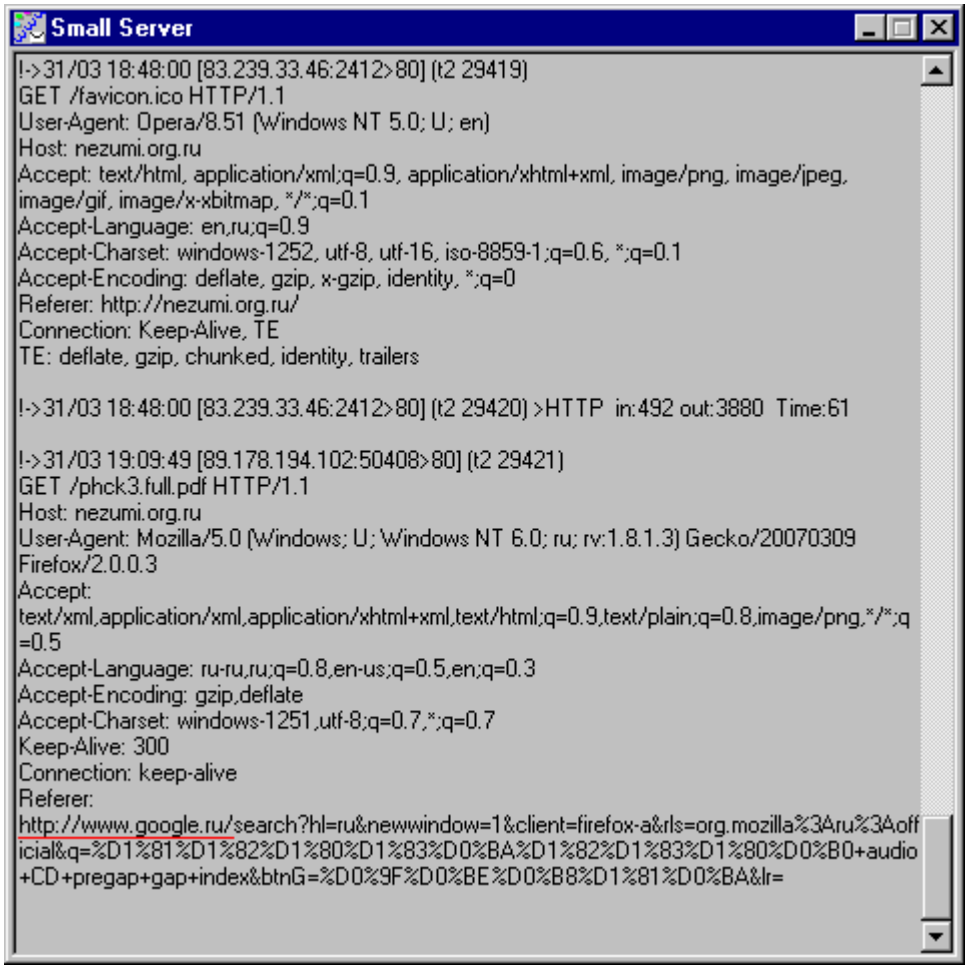

Рисунок 11 консоль Small Http сервера

Ну ладно, зашел человек, так зашел. Ведь не прогонять же!!! А вот вставить содержимое поля referrer в адресную строку Горящего Лиса — сам Хвост велел. Вставляем. И… видим что ***на самом деле*** искал человек (см. рис. 12). А искал он "структура audio CD pregap gap index", причем из трех выданных результатов его удовлетворил только один — мой.

Довольно любопытная информация, не правда ли? Впрочем, остальные поисковые машины страдают той же болезнью, так что Google в своих проблемах не одинок.

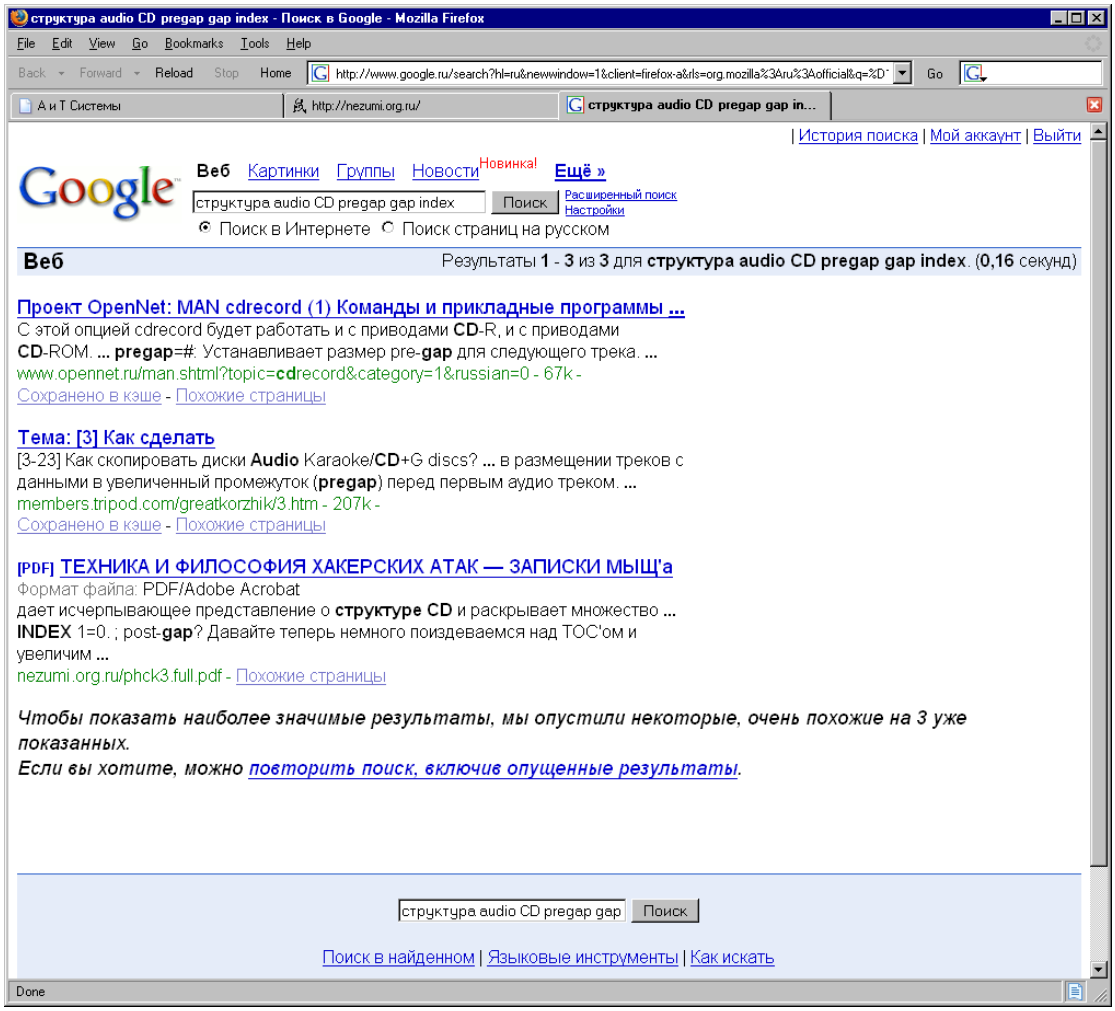

Рисунок 12 перехват чужой поисковой сессии

Да что там поисковые машины! С почтовыми клиентами, основанными на web-интерфейсе, сплошь и рядом наблюдается та же проблема. Устанавливаем у себя web-сервер, отмыливаем жертве ссылку на какой-нибудь интересный файл, если она поведется и кликнет, мы заполучим referrer и, скопировав его в адресную строку своего браузера, сможем войти в текущую сессию, просматривая содержимое почтового ящика жертвы (входящие/исходящие письма), листая адресную книгу и рассылая письма от ее имени. Правда, сменить пароль скорее всего не получиться, равно как и удалить аккаунт, но все-таки… говорить о "безопасности" в таких условиях можно только в саркастическом смысле.

методы борьбы или записки из подполья

Для предотвращения утечки персональной информации на клиентской стороне достаточно использовать Горящего Лиса с установленной панелью Google toolbar и отключенными опциями "PageRank Display" и "Safe Browsing", однако, это не защитит от сайтов, сотрудничающих с Google и для блокирования трафика разумно прибегнуть к персональному брандмауэру, пополнив Black-List еще одной записью: www.google-analytics.com.

С Оперой ситуация значительно сложнее и для обеспечения надлежащего уровня приватности необходимо заблокировать множество IP-адресов, входящих в распределенную сеть Google, постоянно пополняющуюся новыми узлами. Регулярное изучение логов брандмауэра — похоже единственный вариант… как их вычислить по другому мыщъх не знает.

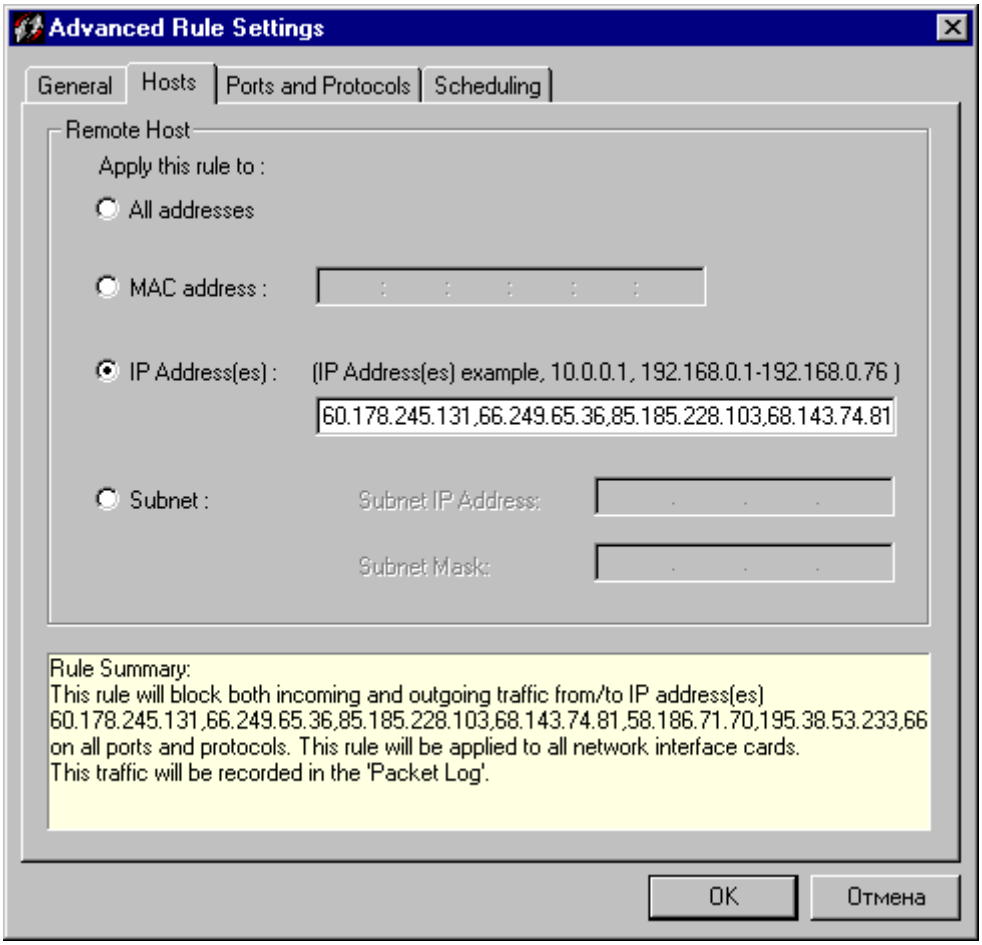

Рисунок 13 блокирование IP-адресов, принадлежащим узлам распределенной сети Google на персональном брандмауэре SyGate Personal Firewall

Администраторам web-серверов рекомендуется блокировать всех посетителей, чье поле User-Agent содержит какое-либо упоминание о Google или создать файл robots.txt, предназначенный специально для поисковых машин и указывающий им, какие файлы можно индексировать, а какие — нет (структура файла описана на http://www.robotstxt.org), впрочем, это достаточно ненадежная защита и поисковые машины могут игнорировать все предписания.

заключение

Доступность исходных кодов еще не гарантирует отсутствие закладок и других компонентов о существовании которых рядовой пользователь не догадывается и мягко говоря, совершенно не нуждается. Но увы… слежка и шпионаж проникают в нашу жизнь и разрушают право на охрану персональной информации, как термиты, подтачивая ее изнутри. Залогом выживания в этом суровом мире становятся знание сетевых протоколов, владение дизассемблером, отладчиков наряду с прочими хакерскими навыками.