УДК 004.056, 004.051

К. И. Будников, **А. В. Курочкин**, **А. А. Лубков**, **А. В. Яковлев**

Институт автоматики и электрометрии СО РАН
пр. Акад. Коптюга, 1, Новосибирск, 630090, Россия

budnikov@iae.nsk.su

МЕТОД ФИЛЬТРАЦИИ HTTP-ПАКЕТОВ НА ОСНОВЕ ПОСТАНАЛИЗА ЗАПРОСОВ К WEB-РЕСУРСУ

Предложен метод фильтрации HTTP-запросов на пакетном уровне с применением постанализа (последующего анализа) прошедшего запроса для управления доступом к web-ресурсу. В отличие от методов, использующих предварительный анализ запросов, в данном случае анализ запроса происходит не перед его отправкой в Интернет к web-серверу, на котором расположен запрашиваемый ресурс, а после отправки, в то время, пока запрос по линиям связи доходит до web-сервера, который формирует ответ, и ответное сообщение возвращается к фильтру. По результатам проверки полученный от web-сервера ответ либо пропускается фильтрующим устройством к пользователю, либо блокируется. Такой подход позволяет уменьшить время ожидания выполнения запроса к ресурсу по сравнению с методами, использующими анализ запроса до его отправки web-серверу.

Ключевые слова: фильтрация HTTP-трафика, анализ сетевых пакетов, регламентирование доступа к web-ресурсу.

Введение

В настоящее время Интернет является крупнейшим хранилищем информации. По зарубежным оценкам компании IDC (США) [1], объем Интернет-контента в 2012 г. составлял около 2,5 млрд Тбайт и удваивался примерно каждые полтора года. Предполагается, что с 2013 по 2020 г. произойдет десятикратное увеличение объема информации в Интернете – с 4,4 до 44 млрд Тбайт.

Развитие Интернета сопровождается появлением большого количества информационных ресурсов, доступ к которым требует ограничения по различным критериям: возрастным, морально-этическим, по требованиям соблюдения безопасности, авторских прав, трудового режима и т. п. При решении данной задачи можно воспользоваться такими методами, как ограничение доступа по IP-адресу, адресу URL, путем изменения запросов к DNS-серверам, использованием прокси-серверов, пакетной фильтрацией. Эти подходы имеют свои достоинства и недостатки [2]. Наиболее сбалансированным по соотношению достоинств и недостатков можно признать способ фильтрации запросов к ресурсу по его адресу URL. Этот метод позволяет осуществить фильтрацию конкретного ресурса. Суть метода заключается в следующем. Фильтрующее устройство перехватывает проходящий через него запрос пользователя, выделяет из него адрес ресурса, к которому происходит обращение, ищет его в списках запрещенных адресов и формирует соответствующие полученному результату действия. Если адрес URL, к которому происходит обращение, не запрещен, то запрос пропускается в Интернет, доходит до сервера с необходимым ресурсом, и сервер возвращает ответ с запрашиваемой информацией. Если доступ к интересующему пользователя ресурсу запрещен, то запрос блокируется фильтром.

Фильтрация по адресу URL может осуществляться как для отдельного устройства доступа в Интернет (компьютер, смартфон, планшет), так и для группы устройств. В пер-

Будников К. И., *Курочкин А. В.*, *Лубков А. А.*, *Яковлев А. В.* Метод фильтрации HTTP-пакетов на основе постанализа запросов к web-ресурсу // Сибирский физический журнал. 2017. Т. 12, № 1. С. 13–18.

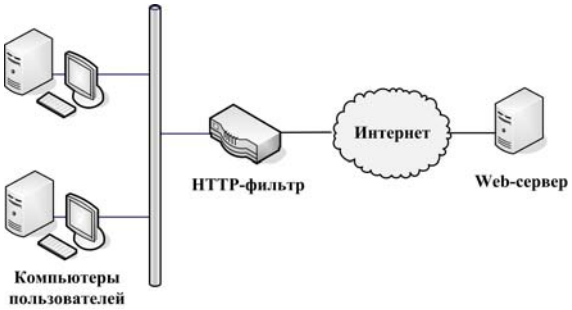

Рис. 1. Схема подключения фильтрующего устройства

вом случае процесс фильтрации осуществляет специально установленная программа, а во втором – фильтрующее устройство, имеющее выход в Интернет, к которому подсоединены компьютеры пользователей (рис. 1). Первый подход предложен, например, в патентах [3; 4]. В качестве примеров второго подхода назовем патенты [5; 6].

Достоинство первого подхода в относительной простоте его реализации исключительно программными средствами. Однако он обладает таким существенным недостатком, как потенциальная возможность пользователя отключить фильтрующую программу по своему усмотрению и таким образом обойти процесс фильтрации.

При реализации подхода второго типа работоспособность фильтрующего устройства не может регулироваться подключенными к нему пользователями. Информацию, необходимую для принятия решения о фильтрации адресов, подобное устройство получает от внешних серверов, называемых серверами фильтрации или репутационными серверами, через сетевые соединения. Для пользователей фильтры прозрачны и представляют собой только линии задержки на пути запроса к интересующему его web-ресурсу. Чем быстрее проходит запрос через фильтрующее устройство, тем менее заметно его присутствие для пользователя и тем большее количество запросов может проходить через фильтр. Алгоритм фильтрации в подобных устройствах (см., например, [5; 6]) предполагает предварительную проверку запроса на входе устройства (рис. 2), и только по ее результатам запрос либо пропускается дальше (рис. 2, *а*), либо блокируется (рис. 2, *б*).

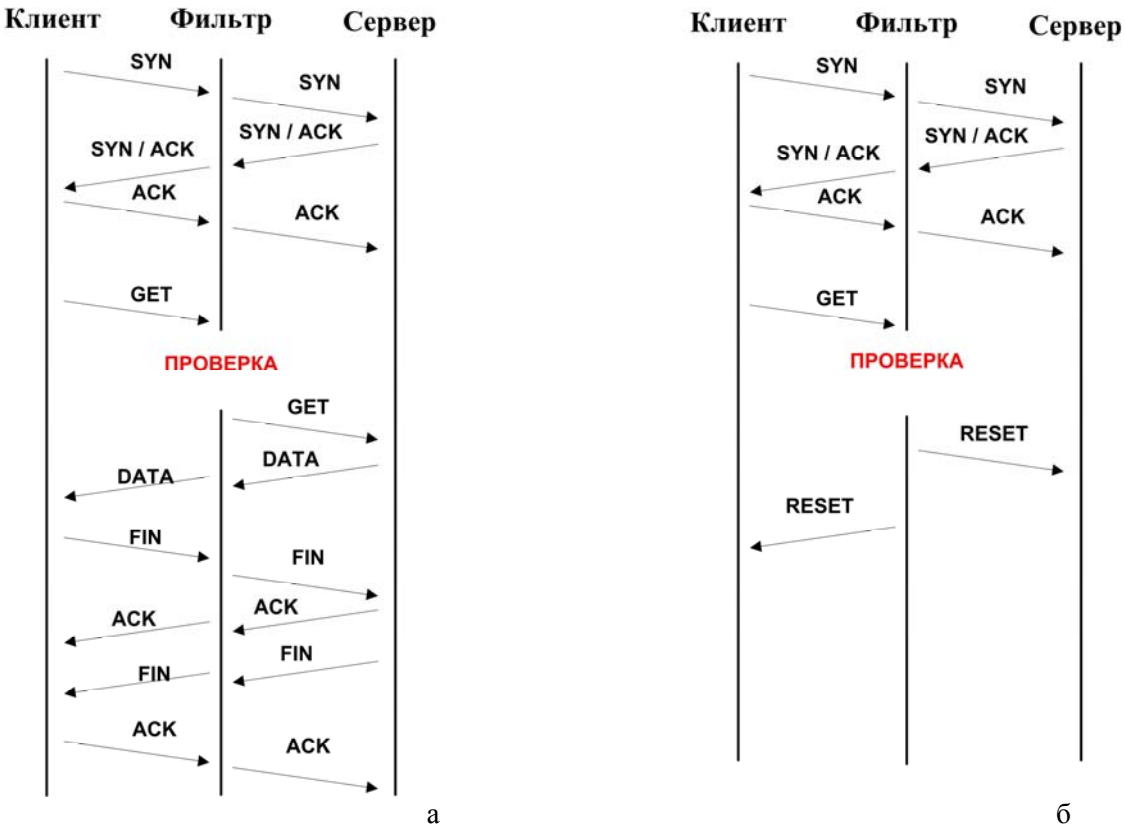

Рис. 2. Временная диаграмма прохождения запроса пользователя через фильтрующее устройство с предварительным анализом запроса к ресурсу. Запрос успешно проходит через фильтр (*а*) или блокируется фильтром путем разрыва TCP-соединения (*б*)

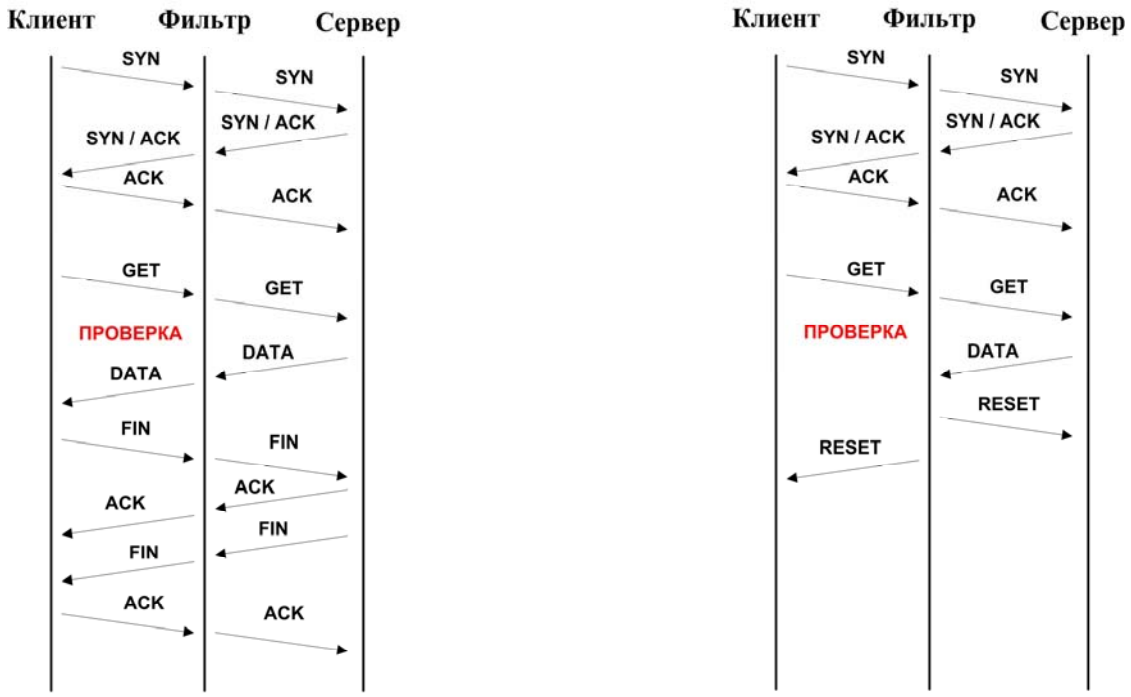

Рис. 3. Временная диаграмма прохождения запроса пользователя через фильтрующее устройство с постанализом запроса к ресурсу. Запрос успешно проходит через фильтр (*а*) или блокируется фильтром путем разрыва TCP-соединения (*б*)

Накладные расходы на проверку составляют временные затраты, связанные с перехватом запроса, выделением из него адреса URL и поиском его в списках запрещенных адресов. На этот период запрос задерживается фильтром. На продолжительность процедуры проверки также существенно влияет длительность запроса к серверу фильтрации, необходимость в котором возникает в случае отсутствия информации об адресе URL в локальных списках запрещенных адресов. Задержка запроса фильтрующим устройством приводит к увеличению времени ожидания ответа от web-сервера, на котором расположен запрашиваемый ресурс. Уменьшение времени задержки прохождения пользовательского запроса через фильтр возможно за счет коррекции алгоритма обработки пакетов, использования метода постанализа запросов к web-ресурсу вместо предварительного анализа. Это позволит обеспечить приемлемое время отклика для большего числа пользователей Интернета, чьи запросы проходят через фильтр, чем при использовании предварительного анализа запросов.

Метод постанализа запросов к web-ресурсу

Предлагаемый метод фильтрации [7] заключается в том, что процессы выполнения запроса и его проверки фильтром происходят параллельно (рис. 3).

В фильтре, использующем постанализ HTTP-запросов, все пакеты, поступающие на вход устройства, в том числе и содержащие запрос пользователя, всегда пропускаются на выход устройства без задержки и изменения, а для анализа запроса создается копия пропущенных пакетов протокола HTTP. В фильтрующих устройствах, использующих предварительный анализ HTTP-запросов, проверка запроса, поступившего на вход фильтра, происходит перед его отправкой в Интернет. Запрос, который во время проверки признан разрешенным, пропускается на выход фильтра и далее к web-серверу с запрашиваемым ресурсом. Фильтр, который применяет пост-анализ HTTP-запросов, проверяет копию поступившего запроса после его отправки в Интернет. Проверка происходит в то время, пока запрос

по линиям связи доходит до web-сервера, на котором расположен запрашиваемый ресурс, формируется ответ со стороны сервера, который возвращается к фильтру, т. е. в режиме постанализа (последующего анализа) прошедшего запроса. По результатам проверки полученный от web-сервера ответ либо пропускается к пользователю (рис. 3, *а*), либо блокируется (рис. 3, *б*).

Модель устройства и алгоритм работы

Представленный метод фильтрации может быть проиллюстрирован на примере работы упрощенной модели пакетного фильтра (рис. 4).

Фильтрующее устройство устанавливается в разрыв между локальной сетью, к которой подсоединены компьютеры пользователей и глобальной сетью Интернет с web-серверами, предоставляющими ресурсы по протоколу HTTP, как показано на рис.1. Модель фильтра состоит из интерфейса сети пользователя (ИСП), интерфейса сети Интернет (ИСИ), двух селекторов (С1 и С2), анализатора-корректора (АК), хранилища текущего состояния контролируемых TCP-сессий (СКС) и хранилища идентификаторов запрещенных ресурсов (ИЗР)

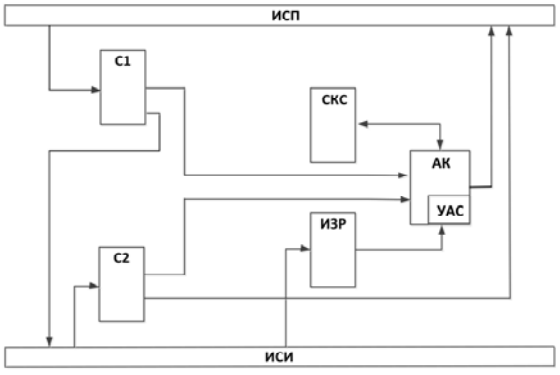

Рис. 4. Модель устройства фильтрации, реализующего метод постанализа запросов к web-ресурсу

Селекторы выделяют пакеты протокола HTTP из общего трафика, причем С1 пропускает весь трафик, поступающий от ИСП, в ИСИ фильтра незамедлительно и без изменений, а копии HTTP-пакетов направляет в АК. С2 пропускает весь трафик, поступающий с ИСИ на ИСП, за исключением пакетов протокола HTTP, которые направляются в АК для проверки. АК, используя информацию из ИЗР, выполняет проверку запросов на право доступа к запрашиваемому ресурсу и при необходимости блокирует получение пользователем ответа от web-сервера с запрещенным контентом.

В АК встроен узел формирования и анализа сессий (УАС), который из сетевых пакетов протокола HTTP, формирует TCP-сессии, в рамках которых пользователями запрашиваются те или иные web-ресурсы, хранит информацию об этих сессиях в СКС и по запросу предоставляет статус запроса: является он разрешенным или нет.

Модель функционирует следующим образом. Поток пакетов с запросом пользователя, на пути к web-серверу достигнув фильтрующего устройства, попадает в интерфейс сети пользователя ИСП, а из него – в первый селектор. С1 отправляет пакеты, составляющие запрос, в интерфейс сети Интернет ИСИ (т. е. к web-серверу), а копии этих пакетов – в анализатор-корректор. В нем формируется TCP-сессия, из запроса выделяется идентификатор ресурса URL и проводится проверка права доступа к этому ресурсу. Таким образом, проверка копии запроса пользователя происходит параллельно с транспортировкой оригинала пользовательского запроса от устройства фильтрации до web-сервера и ответа от web-сервера пользователю до устройства фильтрации.

Ответ от web-сервера, достигнув устройства фильтрации, поступает в интерфейс сети пользователя и далее во второй селектор. С2 отделяет пакеты TCP-сессий протокола HTTP и передает в анализатор-корректор для последующей обработки. УАС, запросив СКС, определяет для пакета соответствующую ему TCP-сессию из списка контролируемых TCP-сессий. Затем проверяется, разрешен ли текущий запрос для этой TCP-сессии. Если запрос разрешен, то пакет посылается в интерфейс сети пользователя без изменений. В противном случае производятся действия, связанные с конкретным алгоритмом блокировки ответа (уничтожение пакета, модификация данных, посылка предупреждения о блокировке и т. п.).

Выигрыш во времени прохождения пользовательского запроса через устройство фильтрации при использовании постанализа по сравнению с предварительным анализом

составляет время, потраченное на определение TCP-сессии для каждого пакета, формирование из пакетов пользовательского запроса, извлечение идентификатора запрашиваемого ресурса URL и проверки запроса на право доступа к запрашиваемому ресурсу по внутренним спискам запрещенных URL.

Компьютерное моделирование устройства фильтрации

Для получения экспериментальной оценки экономии времени при прохождении пользовательского запроса через фильтр при использовании метода постанализа по сравнению с методом предварительного анализа запроса было проведено компьютерное моделирование устройства фильтрации. Для того чтобы исключить влияние сетевой инфраструктуры на работу модели, которое присутствует при стендовых испытаниях (см., например, [8]), интерфейсы сети пользователя и сети Интернет эмулировались программно, и все сетевые потоки данных протекали в памяти моделирующего компьютера.

Результаты моделирования зависят от ряда факторов: мощность используемого компьютера, программная реализация модели фильтра, степень загрузки модели, состав и интенсивность трафика, проходящего через модель устройства фильтрации, и т. п. При моделировании эмулировался процесс непрерывной посылки через фильтр запросов к web-ресурсу и получения ответов через устройство группой пользователей.

Компьютерное моделирование показало уменьшение среднего времени прохождения пользовательского запроса к web-ресурсу через эмулируемое устройство фильтрации, которое работало в режиме постанализа запросов, до 14 % по сравнению с устройством, которое работало в режиме предварительного анализа запросов.

Заключение

Для регулирования пользовательских обращений к информации, размещенной на web-сервере, предложен метод фильтрации, основанный на применении постанализа запросов к web-ресурсу. В отличие от традиционных методов, в фильтрующем устройстве анализ запроса происходит параллельно с его отправкой в Интернет и прохождением по линиям связи до web-сервера, на котором расположен запрашиваемый ресурс, формированием web-сервером ответа, его отправкой пользователю и получением фильтром. По результатам проверки в фильтрующем устройстве пришедший от web-сервера ответ либо пропускается к пользователю, либо блокируется. Такой подход позволяет уменьшить время ожидания пользователем запрашиваемой информации по сравнению с подходом, использующим во время фильтрации предварительный анализ запроса до его отправки к web-серверу.

Список литературы

1. *Turner V.*, *Gantz J. F.*, *Reinsel D.*, *Minton S.* The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things // The EMC Digital Universe Study. EMC Corp., 2014. URL: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf; http://www.slideshare.net/FranciscoCalzado/201404-white-paper-digital-universe-2014 (дата обращения 10.11.2016).

2. *Апетьян С.*, *Ковалев А.*, *Файб А.* Фильтрация контента в Интернете. Анализ мировой практики // Фонд развития гражданского общества. 2013. 22 мая. URL: http://civilfund.ru/Filtraciya_Kontenta_V_Internete_Analiz_Mirovoy_Praktiki.pdf (дата обращения 10.11.2016).

3. *Осипов Г. С.*, *Тихомиров И. А.*, *Соченков И. В.* Способ и система фильтрации вебконтента // Изобретения. Полезные модели: Официальный бюллетень Роспатента. 2012. № 9. Патент № 2446460.

4. *Бейлинсон К. А.*, *Эванс К. А.*, *Фрэверт Г. Дж. В.*, *Тэйлор В. Р.* Фильтрация контента при веб-просмотре // Патент RU 2 336 561 C2. МПК G06F17/30, G06F13/00, H04L12/22. Опубликован 20.10.2008.

5. *Bloch E.*, *Mohan Sh.*, *Pagaku R. R. et al*. Apparatus for monitoring network traffic // Patent US 7849502 B1, Int Cl G06F 15/16 (2006.01), G06F 11/00 (2006.01). Publ. Date: Dec. 7, 2010.

6. *Jai Balasubrahmaniyan*, *Kuntal Daftary*, *Venkateswara Rao Yarlagadda*, *Krishna Kumar*. System and method for URL filtering in a firewall // Patent US 20060064469A1, Int. Cl.G06F 15/16 (2006.01), Publ. Date: Mar. 23, 2006.

7. *Будников К. И.*, *Курочкин А. В.* Способ фильтрации потока HTTP-пакетов на основе

постанализа запросов к Интернет-ресурсу и устройство фильтрации для его реализации // Изобретения. Полезные модели: Официальный бюллетень Роспатента. 2016. № 29. Патент № 2599949.

8. *Будников К. И.*, *Курочкин А. В.*, *Лубков А. А.*, *Яковлев А. В*. Метод экспериментальной оценки датчиков мониторинга электронной почты // Вестн. Новосиб. гос. ун-та. Серия: Физика. 2012. Т. 7, вып. 1. С. 87–93.

Материал поступил в редколлегию 15.11.2016

K. I. Budnikov, **A. V. Kurochkin**, **A. A. Lubkov**, **A. V. Yakovlev**

Institute of Automation and Electrometry SB RAS
1 Acad. Koptyug Ave., *Novosibirsk*, *630090*, *Russian Federation*

budnikov@iae.nsk.su

HTTP PACKETS FILTRATION METHOD BASED ON POST-ANALYSIS OF REQUESTS TO WEB-RESOURCE

The paper describes an HTTP-request filtering method at packet level by using post-analysis of the passed request to control access to a web-resource. Unlike methods using preliminary analysis of requests, in this method the request is analyzed not before it is sent to the Internet to the web-server on which the requested resource is placed, but after, at the time while the request over communication lines comes to the web-server, which generates a response, and the response message comes back to the filter. Depending on the check results the response received from the web-server by filter device is passed to the user or locked. The approach of such kind reduces the time to wait for the answer on request to the resource in comparison to methods that use preliminary analysis of the request before it is sent to web-server.

Keywords: HTTP-traffic filtration, network packets analyzing, regulation of access to web-resource.