УПРАВЛЕНИЕ ДОСТУПОМ К WEB-РЕСУРСУ НА ОСНОВЕ ПОСТ-АНАЛИЗА HTTP-ЗАПРОСОВ

К.И. Будников, А.В. Курочкин, А.А. Лубков, А.В. Яковлев
Институт автоматики и электрометрии Сибирского отделения РАН, Новосибирск, Россия

Предложен метод управления доступом к web-ресурсу путем фильтрации HTTP-запросов на пакетном уровне с применение пост-анализа (последующего анализа) HTTP-запросов, прошедших через устройство фильтрации. В данном методе анализ запроса происходит не перед его отправкой в Интернет, а после отправки, в то время, пока запрос по линиям связи Интернет доходит до web-сервера, на котором расположен запрашиваемый ресурс, web-сервер формирует ответ, и этот ответ доходит обратно до фильтра. 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 предварительный анализ запроса до его отправки web-серверу.

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

Введение

В последние годы развитие Интернет сопровождается появлением большого количества информационных ресурсов, доступ к которым требует ограничения по различным критериям: возрастным, морально-этическим, требованиям соблюдения безопасности, авторских прав, трудового режима и т.п. Данная задача может решаться 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 [1]. 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 перехвате фильтром запроса пользователя, выделении из него адреса ресурса, поиска его в списках запрещенных адресов и формировании соответствующих полученному результату действий. Если адрес не запрещен, то запрос пропускается в Интернет, доходит до сервера с необходимым ресурсом, который возвращает ответ с запрашиваемой информацией. Если доступ к интересующему пользователя ресурсу запрещен, то запрос блокируется.

Фильтрация по адресу URL может осуществляться как для отдельного устройства доступа в Интернет (компьютер, смартфон, планшет), так и для группы устройств. В первом случае процесс фильтрации осуществляет специально установленная программа, а во втором случае - фильтрующее устройство, имеющее выход в Интернет, к которому подсоединены компьютеры пользователей (рис.1). Первый подход предложен, например, в патентах [2,3]. В качестве примеров второго подхода служат патенты [4,5].

При относительной простоте реализации первого подхода (может быть реализован программными средствами), он обладает таким существенным недостатком, как потенциальная возможность пользователя отключить фильтрующую программу по своему усмотрению и таким образом обойти процесс фильтрации.

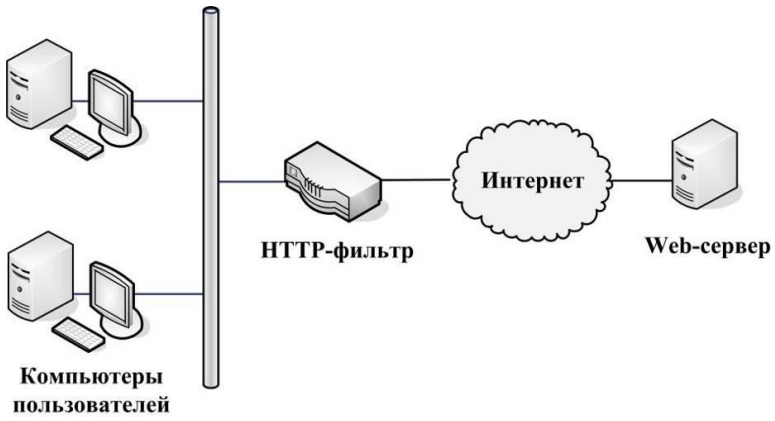

Рис. 1. Схема подключения фильтрующего устройства

Работоспособность устройств, реализующих подход второго типа, не может регулироваться подключенными к ним пользователями. Информацию, необходимую для принятия решения о фильтрации адресов такие устройства получают 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 пользователя к интересующему его ресурсу. Чем быстрее проходит запрос через фильтрующее устройство, тем менее заметно его присутствие для пользователя и большее количество запросов может проходить через фильтр.

Алгоритм фильтрации в подобных устройствах, (см. например [4,5]) предполагает предварительную проверку запроса на входе устройства, и только по ее результатам запрос либо пропускается дальше, либо блокируется. Сама проверка занимает время, связанное с перехватом запроса, выделением из него адреса URL и поиском его в списках запрещенных адресов. На это время 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. Уменьшение времени задержки прохождения пользовательского запроса возможно за счет коррекции алгоритма обработки пакетов, использовании метода пост-анализа запросов к web-ресурсу вместо предварительного анализа. Это позволит обеспечить приемлемое время отклика для большего числа пользователей Интернет, чьи запросы проходят через фильтр.

Метод пост-анализа запросов к web-ресурсу

Предлагаемый метод фильтрации заключается в нижеследующем. В фильтре, использующем пост-анализ 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. Проверка HTTP-запроса происходит не перед его отправкой в Интернет, как это делается в фильтрующих устройствах, использующих предварительный анализ HTTP-запросов, а после отправки запроса в Интернет, в то время, пока запрос по линиям связи доходит до web-сервера, на котором расположен запрашиваемый ресурс, формиру-

ется ответ со стороны сервера, и он доходит обратно до фильтра, т.е. в режиме пост-анализа (последующего анализа) запроса. По результатам проверки полученный от web-сервера ответ либо пропускается к пользователю, либо блокируется.

Модель устройства и алгоритм его работы

Представленный метод фильтрации может быть проиллюстрирован на примере работы упрощенной модели пакетного фильтра, представленной на рис. 2. 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).

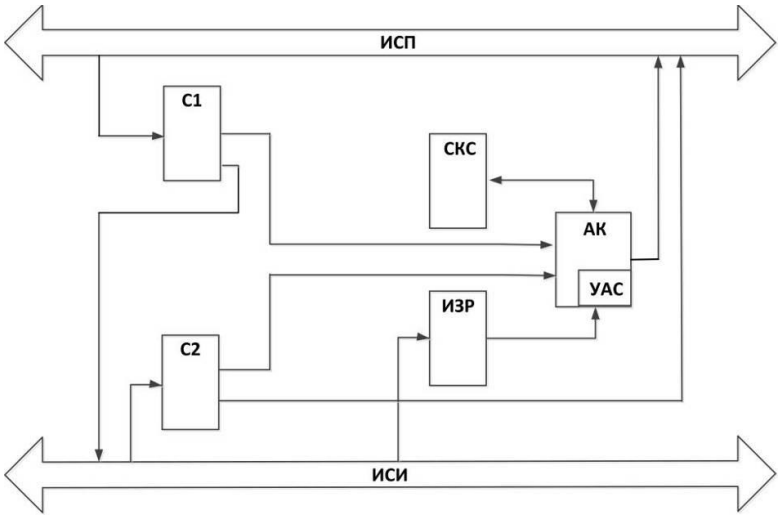

Рис. 2. Модель устройства фильтрации, реализующего метод пост-анализа запросов к web-ресурсу

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 ли запрос разрешенным или нет.

Модель функционирует следующим образом. Поток пакетов с запросом пользователя на пути к web-серверу достигнув фильтрующее устройство попадает в интерфейс сети пользователя ИСП, а из него в первый селектор. С1 отправляет пакеты, составляющие за-

прос, в интерфейс сети Интернет ИСИ (то есть, к web-серверу), а копии этих пакетов – в анализатор-корректор, где 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. Узлом формирования и анализа TCP-сессий для пакета определяется соответствующая TCP-сессия в списке контролируемых TCP-сессий в узле хранения текущего состояния контролируемых TCP-сессий и 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 пост-анализа 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 временного выигрыша при прохождении пользовательского запроса через фильтр при использовании метода пост-анализа по сравнению с методом предварительного анализа запроса было проведено компьютерное моделирование устройства фильтрации. Для того, чтобы исключить влияние сетевой инфраструктуры на работу модели, интерфейсы сети пользователя и сети Интернет эмулировались программно, и все сетевые потоки данных протекали в памяти моделирующего компьютера.

Выигрыш во времени при моделировании зависит от ряда факторов, включая мощность используемого компьютера, программную реализация модели фильтра, степень загрузки модели, состав и интенсивность трафика, проходящего через модель устройства фильтрации и т.п. В процессе моделирования эмулировался процесс непрерывной посылки через фильтр запросов к web-ресурсу и получения ответов через устройство группой пользователей.

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

Заключение

Для фильтрации запросов к web-ресурсу, предложен метод, основанный на применении пост-анализа запросов к web-ресурсу. В отличие от традиционных методов, анализ запроса в фильтрующем устройстве происходит не перед отправкой запроса в Интернет, а после, в то время, пока запрос по линиям связи доходит до web-сервера, на котором расположен запрашиваемый ресурс, формируется ответ со стороны web-сервера, и этот ответ доходит обратно до фильтра. По результатам проверки полученный от web-сервера ответ либо пропускается к пользователю, либо блокируется. Такой подход позволяет уменьшить время ожидания ответа на запрос к ресурсу по сравнению с подходом, использующим предварительный анализ запроса до его отправки к web-серверу.

Литература

1. Апетьян, Станислав «Фильтрация контента в Интернете. Анализ мировой практики»/ Андрей Ковалев, Александр Файб // Фонд развития гражданского общества, 22 мая, 2013. [Электронный ресурс] - URL: http://civilfund.ru/Filtraciya_Kontenta_V_Internete_Analiz_Mirovoy_Praktiki.pdf (дата обращения: 31.03.2016)
2. Осипов, Г.С. Способ и система фильтрации веб-контента / И.А.Тихомиров, И.В.Соченков // патент RU 2446460 C1. МПК G06F 21/20 (2006.01), опубликован 27.03.2012г. Бюл. № 9
3. Бейлинсон, К.А. Фильтрация контента при веб-просмотре /К.А. Эванс, Г.Дж. В. Фрэверт, В.Р. Тэйлор...// патент RU 2 336 561 C2. МПК G06F17/30, G06F13/00, H04L12/22, опубликован 20.10.2008 г.
4. Bloch, Eric Apparatus for monitoring network traffic/ Shalabh Mohan, Rajendraprasad R. Pagaku, et al.// patent US 7849502 B1, Int Cl G06F 15/16 (2006.01), G06F 11/00 (2006.01), Pub. Date: Dec. 7, 2010.
5. Balasubrahmaniyan, Jai,. SYSTEM AND METHOD FOR URL FILTERING IN A FIREWALL /Kuntal Daftary, Venkateswara Rao Yarlagadda, Krishna Kumar // patent US 20060064469A1, Int. Cl.G06F 15/16 (2006.01), Pub. Date:Mar. 23, 2006.