УДК 004.056, 004.051
DOI 10.25205/2541-9447-2018-13-1-5-12

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

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

*budnikov@iae.nsk.su*

# МЕТОДИКА ЭКСПЕРИМЕНТАЛЬНОГО ИССЛЕДОВАНИЯ HTTP-ФИЛЬТРОВ *

Фильтрация запросов пользователей к Интернет-ресурсам позволяет осуществлять регулирование доступа к затребованной ими информации. В статье представлена методика стендового исследования характеристик устройств фильтрации по протоколу HTTP, которая включает схему и процедуру проведения испытаний, способ получения и оценки свойств изучаемого прибора. Рассмотрена двухканальная симметричная компьютерная модель фильтра HTTP-запросов, функционирующая на пакетном уровне, использующая метод ограничения доступа к Интернет-ресурсу по его адресу URL. Приведены результаты экспериментальных испытаний модели по предложенной методике.

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

## Введение

Рост объема интернет-контента и наличие большого количества информационных интернет-ресурсов, доступ к которым нуждается в ограничении по ряду критериев, включая возрастные и морально-этические, необходимость соблюдения безопасности, защиты авторских прав, трудового режима и т. п. требует совершенствования средств и методов [1; 2] обеспечения селективного запрета на доступ к информации в сети. В настоящее время эти методы включают: ограничение доступа по IP-адресу, по адресу URL, изменение запросов к DNS-серверам, использование прокси-серверов, пакетной фильтрации. Данные подходы имеют как достоинства, так и недостатки. Наиболее сбалансированным в этом отношении можно признать способ фильтрации запросов к ресурсу по его адресу URL. Этот метод позволяет производить фильтрацию конкретного интернет-ресурса, не блокируя остальные, расположенные на том же сервере. Для отдельного устройства доступа в Интернет (компьютер, смартфон, планшет) процесс фильтрации осуществляет специально установленная программа [3; 4], а для группы устройств – фильтрующий прибор, имеющий выход в Интернет, к которому они подсоединены [5; 6].

Устройство, осуществляющее фильтрацию по URL, перехватывает проходящий через него запрос пользователя, выделяет из него адрес ресурса, к которому происходит обращение. Далее в зависимости от встроенного алгоритма осуществляется поиск этого адреса в списках ресурсов, которые

* Работа выполнена в рамках НИР госзадания (проект IV.36.1.3 № 031920160009).

*Будников К. И., Курочкин А. В.* Методика экспериментального исследования HTTP-фильтров // Сибирский физический журнал. 2018. Т. 13, № 1. С. 5–12.

запрещены или наоборот разрешены. Если адрес URL, к которому происходит обращение, разрешен, 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.

Фильтр, установленный на пути запроса пользователя к web-ресурсу и ответа от web-сервера, создает задержку при передаче пакетов. Исследование его свойств с целью определения и уменьшения издержек от процесса фильтрации – одна из основных задач моделирования при разработке подобного устройства. Существует ряд методов экспериментальных исследований брандмауэров, которые осуществляют пакетную фильтрацию, например RFC3511 [7]. В основном они предназначены для исследования выполнения функций тестируемым устройством при прохождении через него пакетов в разных режимах передачи, моделируя ту или иную ситуацию, складывающуюся при работе устройства. По отношению к HTTP-фильтру набор тестов для брандмауэра, с одной стороны, избыточен, а с другой – в части, касающейся прохождения пакетов HTTP, недостаточен. С этим связана необходимость разработки методов экспериментального исследования непосредственно HTTP-фильтров.

## Методика исследования устройств HTTP-фильтрации

Разработанная методика включает в себя: схему стенда, процедуру тестирования и оценки полученных данных, необходимое программное обеспечение.

Стенд представляет собой выделенную локальную сеть, объединяющую компьютеры – эмуляторы web-клиента и web-сервера, а также помещенное в разрыв между ними фильтрующее устройство (рис. 1).

Во время испытаний моделируется процесс обращения к интернет-ресурсу по протоколу HTTP. Для этого на компьютер-эмулятор web-клиента установлен программный модуль (К), посылающий запросы компьютеру с программным эмулятором web-сервера (С), который формирует ответы на полученные запросы (рис. 2). HTTP-фильтр (Ф) регламентирует доступ к ресурсам сервера в соответствии с правилами фильтрации. Для определения параметров тестового трафика и ряда характеристик устройства фильтрации на эмуляторе web-клиента размещается программный модуль измерения (И).

Созданный для испытания фильтрующих устройств и исследования их возможностей программный инструментарий позволяет моделировать процесс обращения пользователей к web-серверу, создавая потоки запросов с различной интенсивностью и размерами передаваемых в ответ данных. С целью изучения функционирования модели фильтра в условиях, приближенных к естественным, в которых кроме основного трафика по протоколу HTTP (порт 80) присутствует обмен данными и по другим протоколам, сервер и клиент могут создавать дополнительный «нецелевой» трафик. В этом качестве выступают те же самые сессии HTTP, которые используются для генерации основного потока, только они пересылаются по порту, отличному от 80. Такой подход позволяет относительно точно определять пропорции между основным и дополнительным потоками в общем трафике. Подобный метод применялся при исследовании датчика мониторинга электронной почты [8–10] с использованием сессий почтовых протоколов [11–13].

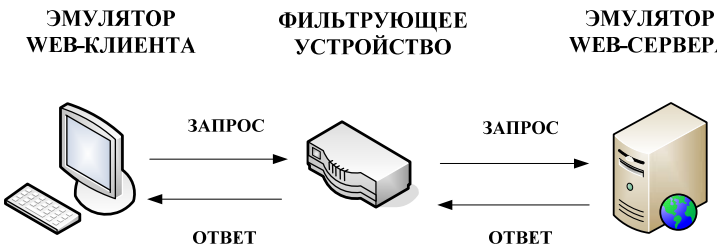


*Рис. 1.* Схема стенда для проведения испытаний фильтрующего устройства

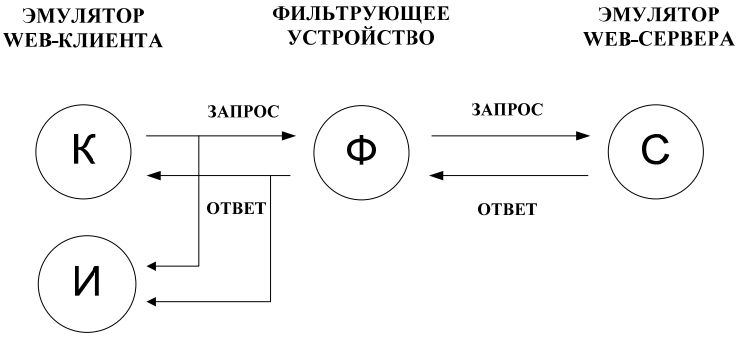


*Рис. 2.* Схема взаимодействия программных модулей стенда для проведения испытаний фильтрующего устройства

Перед проведением испытаний фильтра методика тестирования предполагает калибровку стенда – предварительное измерение его характеристик в отсутствие исследуемого устройства. Впоследствии, уже в процессе исследования характеристик фильтра, это позволяет учесть параметры стенда.

Для анализа результатов и оценки исследуемого устройства модуль измерения предоставляет следующие данные: средняя интенсивность прошедшего через фильтр тестового сетевого трафика; средняя интенсивность HTTP-трафика; средняя интенсивность нецелевого трафика; средняя интенсивность HTTP-запросов; среднее время ожидания ответа web-сервера, среднее время выполнения запроса.

Средняя интенсивность прошедшего через фильтр тестового сетевого трафика $\overline{\lambda}_{\text{тр}}$ (бит/с) рассчитывается по формуле

$$\overline{\lambda}_{\text{тр}} = \frac{\sum_{i=1}^{N} S_i}{T_{\text{эксп}}} \times 8,$$

где $S_i$ – размер *i*-го кадра Ethernet в байтах, включая его данные и служебную информацию, содержащуюся в кадре, а также его преамбулу;

$N$ – количество кадров, прошедших через фильтр;

$T_{\text{эксп}}$ – время проведения эксперимента, вычисляемое как временной интервал между посылкой пакета SYN при установлении первой сессии для HTTP-обмена и получением пакета FIN, означающим окончание последней сессии HTTP-обмена.

Аналогично рассчитываются средние интенсивности HTTP- и нецелевого трафика. Учитываются соответственно HTTP- либо не HTTP-кадры.

Средняя интенсивность HTTP-запросов $\overline{\lambda}_{\text{запр}}$ (1/с) вычисляется по формуле

$$\overline{\lambda}_{\text{запр}} = \frac{N_{http}}{T_{\text{эксп}}}$$

как отношение количества всех HTTP-запросов ко времени проведения эксперимента.

Время ожидания ответа определяется как временной интервал между посылкой пакета с HTTP-запросом и получением первого пакета с ответом от эмулятора web-сервера. Среднее время ожидания ответа $\overline{T}_{\text{ож}}$ определяется по формуле

$$\overline{T}_{\text{ож}} = \frac{\sum_{i=1}^{N} T_i}{N},$$

где $T_i$ – время ожидания ответа, а $N$ – количество сделанных в процессе эксперимента запросов.

Время выполнения HTTP-запроса определяется как временной интервал между посылкой пакета SYN устанавливаемой для HTTP-обмена сессии TCP и получением пакета FIN, означающим окончание этой сессии. Среднее время выполнения HTTP-запроса $\overline{T}_{\text{вып}}$ определяется по формуле

$$\overline{T}_{\text{вып}} \frac{\sum_{i=1}^{N} T_i}{N},$$

где $T_i$ – время выполнения HTTP-запроса, а $N$ – количество сделанных в процессе эксперимента запросов.

Изложенный подход близок к методологии исследования производительности брандмауэров в той части, которая касается протокола HTTP и приведена в RFC3511 (п. 5.6–5.8). Среди основных отличий – схема испытательного стенда, измеряемые характеристики, типы трафика, проходящего через устройство.

В настоящей публикации для сравнительного анализа характеристик фильтра используется время ожидания ответа, которое является одной из важных характеристик информационной системы. Как показали проведенные исследования [14–16], время ответа информационной системы на запрос пользователя в диалоге имеет 4 градации:

1) менее 0,1 с – мгновенная реакция, пользователь не замечает задержку;

2) 0,1–1 с – пользователь может заметить задержку, но не испытывает неудовлетворения;

3) 1–10 с – пользователь может ожидать ответ системы;

4) более 10 с – пользователь переключается на другой диалог.

Исходя из этого необходимо, чтобы время ожидания ответа не превышало 10 с.

В качестве иллюстрации применения изложенной методики представлено исследование симметричной модели фильтрующего устройства.

## Модель устройства фильтрации

Для исследования алгоритмов фильтрации с целью определения издержек и нахождения возможных путей их уменьшения при прохождении пакетов через фильтрующее устройство была создана компьютерная модель [17], которая имеет две симметричные точки подключения и устанавливается в разрыв соединения между сетью клиента и сетью сервера.

Модель состоит из двух равнозначных каналов, которые обеспечивают прохождение через устройство пакетного трафика и его фильтрацию. Каждый канал содержит следующие модули (рис. 3): чтения сетевых пакетов (МЧП), сортировки (МСП) и передачи (МПП) пакетов. Центральный модуль модели – анализатор пакетов (МАП), общий для обоих каналов.

Как показано на рис. 3, клиентская подсеть подключена к входу первого (левого) канала, а серверная – к входу второго (правого). В процессе фильтрации клиентские пакеты считываются МЧП1 и попадают в МСП1, который выделяет пакеты протокола HTTP и передает их в МАП. Остальные пакеты поступают в МПП1. МАП анализирует пакеты и формирует запросы клиента, после чего отправляет пакеты в МПП1. Пакеты, отсылаемые сервером клиенту, проходят по второму каналу: от МЧП2 через МСП2 и МАП в МПП2. Модули МПП передают поступившие к ним пакеты на выход соответствующих каналов. МАП проверяет разрешение для каждого пользовательского запроса и выполняет операции по предотвращению доступа к запрещенному контенту.

## Проведение стендовых испытаний модели

Результаты, полученные в процессе испытаний, зависят от используемого в них оборудования. В экспериментах, результаты которых представлены ниже, в качестве фильтрующего устройства служил компьютер с процессором Intel Core i7 870 2,93 ГГц, 4 Гб оперативной памяти, под управлением ОС Windows 64-битной версии. Испытания

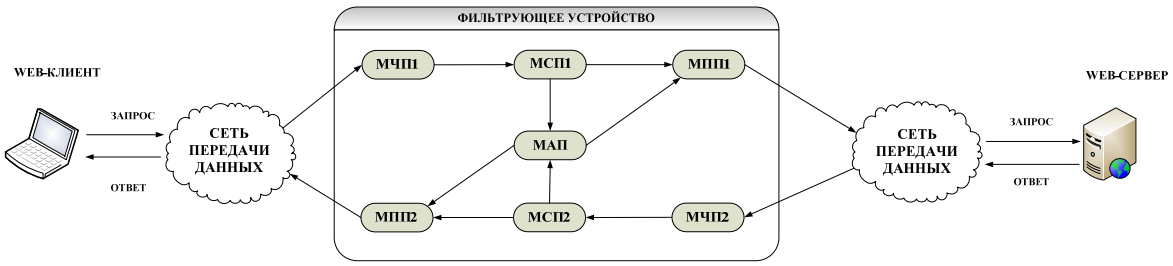


*Рис. 3.* Модель HTTP-фильтра

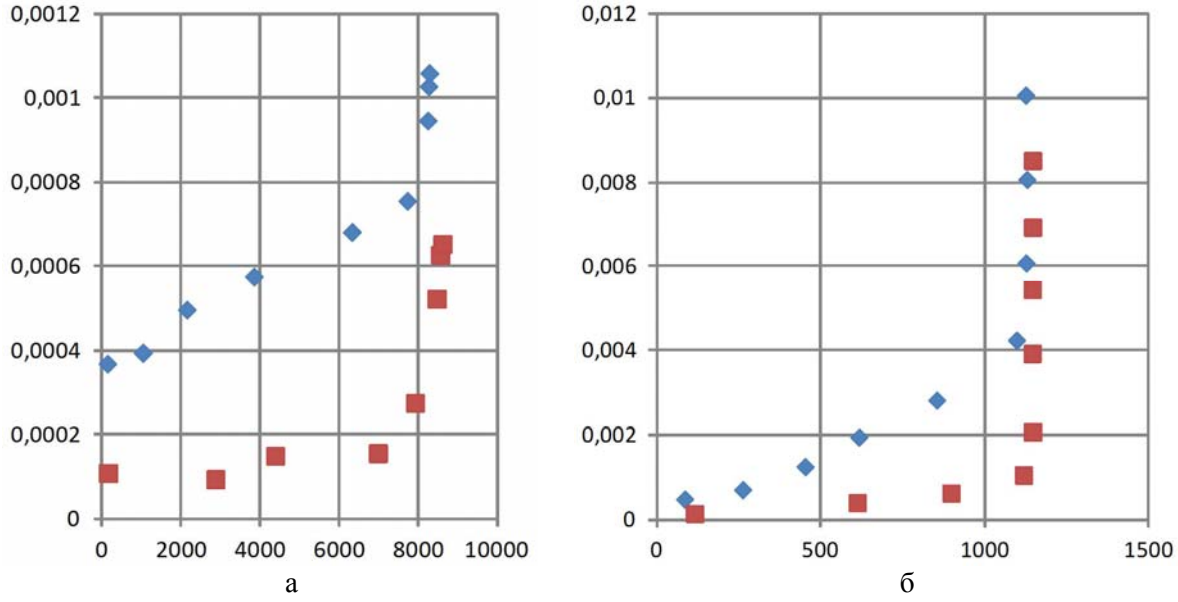


*Рис. 4.* Соотношение отфильтрованного ( ◆ ) и нефильтрованного ( ■ ) трафиков для ответов с размером 1 Кб (*а*) и 100 Кб (*б*)

проводились для сетевых соединений со скоростью передачи данных до 1 Гбит/с. На свои запросы пользователь получал ответы размером от 1 до 100 Кб (рис. 4). В приведенных ниже графиках на одну точку усредненных данных, полученную при испытаниях как результат эксперимента, приходится от 100 000 до 4 200 000 посылок запросов к эмулятору web-сервера и получений ответов от него. Доверительный интервал для доверительной вероятности 95 % колеблется в диапазоне 0,02–0,2 %.

На рис. 4 представлено соотношение трафика, прошедшего через фильтр (отфильтрованного) и трафика, который использовался для измерения характеристик стенда в отсутствие исследуемого устройства (нефильтрованного) при размере ответа 1 и 100 Кб. На каждом графике представлены зависимости среднего времени ожидания запроса, выраженного в секундах (вертикальная ось), от количества запросов к эмулятору web-сервера, проходящих через фильтр в секунду (горизонтальная ось).

Соотношение графиков позволяет выяснить, на какую величину фильтр изменяет пропускную способность моделируемой системы. Например, в приведенных графиках пропускная способность моделируемой системы уменьшается фильтром от 2 до 3 % в зависимости от размера ответа на запрос от эмулятора web-сервера. Время прохождения запросов и ответов через фильтр для исследуемой модели при реализации на использованном оборудовании в сравнении с отсутствием фильтрации увеличилось от 2,5 до 4,6 раза при размере ответа 1 Кб и от 3,4 до 4 раз при размере ответа 100 Кб. Однако максимальное среднее время ожидания ответа не превысило 10 мс, а среднее время выполнения запроса – 88 мс. Таким образом, для смоделированной информационной системы исследуемый вариант фильтра ухудшает ее временные характеристики, но время ожидания ответа находится в приемлемых для пользователя пределах.

Для изучения функционирования модели фильтра в условиях, приближенных к естественным, в которых кроме основного трафика по протоколу HTTP присутствует обмен данными и по другим протоколам, проведены испытания фильтрации потоков с разным процентным содержанием пакетов, содержащих данные протокола HTTP. Графики представлены на рис. 5.

На каждом графике представлены зависимости среднего времени ожидания ответа на запрос 50 и 100 % HTTP-трафика, выраженного в секундах (вертикальная ось), от количества запросов к эмулятору web-сер-

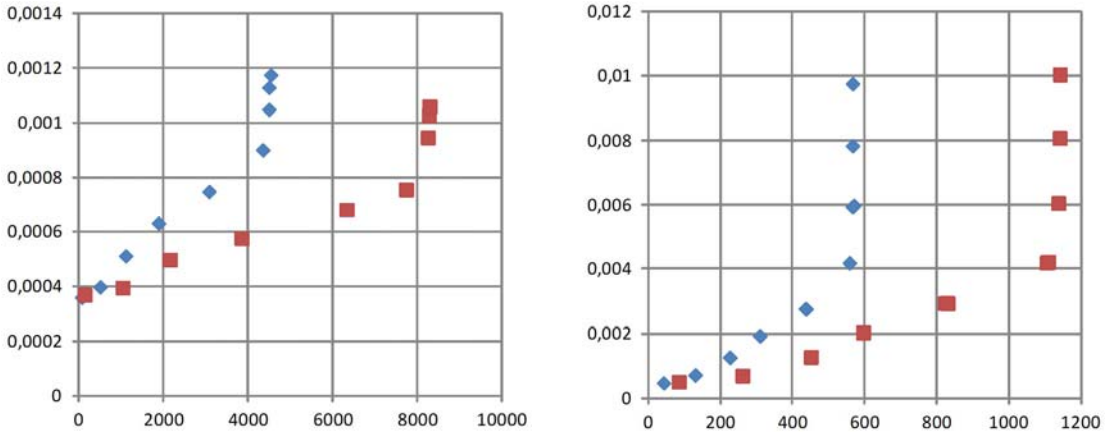


*Рис. 5.* Фильтрация запросов с 50 % (◆) и 100 % (■) содержанием HTTP-трафика при ответе от web-сервера размером 1 Кб (*а*) и 100 Кб (*б*)

вера, проходящих через фильтр в секунду (горизонтальная ось). Из приведенных зависимостей видно, что при трафике с более низким содержанием пакетов протокола HTTP пропускная способность фильтра ниже, а времени, необходимого для фильтрации одного и того же количества запросов и ответов на них, требуется больше, чем при трафике с более высоким содержанием пакетов HTTP.

Так, пропускная способность исследуемой модели фильтра составила 8 278 запросов в секунду при пропуске 100-процентного трафика протокола HTTP с размером ответа 1 Кб, 4 529 запросов в секунду при пропуске 50-процентного трафика протокола HTTP с размером ответа 1 Кб, 1 141 запрос в секунду при пропуске 100-процентного трафика протокола HTTP с размером ответа 100 Кб и 567 запросов в секунду при пропуске 50-процентного трафика протокола HTTP с размером ответа 100 Кб.

## Заключение

В статье представлена методика стендового исследования характеристик HTTP-фильтров, которая включает схему и процедуру проведения испытаний, способ получения и оценки свойств изучаемого устройства. Рассмотрена симметричная двухканальная компьютерная модель устройства фильтрации HTTP-запросов на пакетном уровне, использующая метод фильтрации по адресу URL. Приведены результаты ее экспериментальных исследований по представленной методике, которые показали, что чем выше процент содержания пакетов протокола HTTP в общем потоке, а также чем короче размер ответа от web-сервера, тем выше пропускная способность устройства (выраженная в количестве запросов в секунду) и меньше требуется времени, необходимого для фильтрации одного и того же количества запросов.

## Список литературы

1. *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 20.06.2016).

2. *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.

3. *Осипов Г. С.*, *Тихомиров И. А.*, *Соченков И. В.* Способ и система фильтрации веб-контента // Изобретения. Полезные модели: Официальный бюллетень Роспатента. 2012. № 9. Патент на изобретение № 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. *Eric Bloch*, *Shalabh Mohan*, *Rajendraprasad R. Pagaku et al*. Apparatus for monitoring network traffic // Patent US 7849502 B1, Int Cl G06F 15/16 (2006.01), G06F 11/00 (2006.01), Pub. 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), Pub. Date: Mar. 23, 2006.

7. *Hickman B.*, *Newman D.*, *Tadjudin S.*, *Martin T.* Benchmarking Methodology for Firewall Performance // Network Working Group, Request for Comments: 3511. April 2003. URL: https://tools.ietf.org/html/rfc3511.

8. *Budnikov K. I.*, *Kurochkin A. V.*, *Lylov S. A.* Win32 Based Sensor for Email Auditing // Proc. of the 1st IEEE Region 8 International Conference on «Computational Technologies in Electrical and Electronics Engineering» SIBIRCON-2008. Novosibirsk, 2008. P. 286–287.

9. *Будников К. И.*, *Клисторин И. Ф.*, *Курочкин А. В.*, *Лылов С. А*. Датчик удаленного мониторинга электронной почты // Датчики и системы. 2008. № 9. С. 35–37.

10. *Будников К. И.*, *Клисторин И. Ф.*, *Курочкин А. В.*, *Лылов С. А*. Структурно-функциональная модель интеллектуального датчика мониторинга сетевого трафика // Вестн. компьютерных и информационных технологий. 2011. № 3. С. 51–55.

11. *Будников К. И.*, *Клисторин И. Ф.*, *Курочкин А. В.* Исследование многопоточной модели линейного интеллектуального датчика мониторинга электронной почты на платформе Win32 // Автометрия. 2010. Т. 46, № 5. С. 124–131.

12. *Будников К. И.*, *Курочкин А. В.*, *Лубков А. А.*, *Яковлев А. В.* Метод экспериментальной оценки датчиков мониторинга электронной почты. // Вестн. НГУ. Серия: Физика. 2012, Т. 7, № 1. С. 87–93.

13. *Будников К. И.*, *Курочкин А. В.*, *Лубков А. А.*, *Яковлев А. В.* Синтетический тест TRANSMAIL для оценки датчиков мониторинга электронной почты // Вестн. компьютерных и информационных технологий. 2014. № 5. С. 50–56.

14. *Nielsen J.* Usability Engineering. Cambridge, MA: Academic Press, Inc., 1993. URL: https://www.nngroup.com/articles/response-times-3-important-limits/

15. *Miller R. B.* Response time in man-computer conversational transactions // Proc. AFIPS Fall Joint Computer Conference. 1968. Vol. 33. P. 267–277.

16. *Card S. K.*, *Robertson G. G.*, *Mackinlay J. D.* The information visualizer: An information workspace // Proc. ACM CHI'91 Conf. New Orleans, LA, 1991. P. 181–188.

17. *Budnikov K. I.*, *Kurochkin A. V.*, *Lubkov A. A.*, *Yakovlev A. V*. Experimental Study of Symmetric Computer Model of Http-Filter. // Proc. of the 2nd Russian-Pacific Conference on Computer Technology and Applications RPC 2017. Vladivostok, Russia, 2017.

*Материал поступил в редколлегию 19.12.2017*

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

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

*budnikov@iae.nsk.su*

## METHOD OF EXPERIMENTAL RESEARCH OF HTTP FILTERS

Filtering of user requests to Internet resources allows regulating access to the information requested. The paper presents a technique for bench testing of characteristics of the HTTP protocol

filtering device, which includes the scheme and procedure of testing, the method of obtaining and evaluating the properties of the device under study. A two-channel symmetric computer model of the HTTP request filter is considered. It operates at the packet level, using the method of restricting access to the Internet resource by its URL. The results of experimental tests of the model according to the proposed technique are presented.

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

*For citation*:

Budnikov K. I., Kurochkin A. V. Method of Experimental Research of HTTP Filters. *Siberian Journal of Physics*, 2018, vol. 13, no. 1, p. 5–12. (In Russ.)

DOI 10.25205/2541-9447-2018-13-1-5-12