УДК 004.056

ИССЛЕДОВАНИЕ МНОГОПОТОЧНОЙ МОДЕЛИ ЛИНЕЙНОГО ИНТЕЛЛЕКТУАЛЬНОГО ДАТЧИКА МОНИТОРИНГА ЭЛЕКТРОННОЙ ПОЧТЫ НА ПЛАТФОРМЕ Win32

К. И. Будников, И. Ф. Клисторин, А. В. Курочкин

Учреждение Российской академии наук
Институт автоматики и электрометрии Сибирского отделения РАН,
630090, г. Новосибирск, просп. Академика Коптюга, 1
E-mail: budnikov@iae.nsk.su

Рассмотрено моделирование линейного интеллектуального датчика мониторинга электронной почты на платформе Win32 в среде Windows XP. Программа-эмулятор прибора считывает пакеты с линии связи, обрабатывает их, извлекает почтовую информацию и пересылает её устройству управления. Исследованы особенности и ограничения платформы Win32 для создания приборов данного типа при использовании многопоточного механизма в эмулирующей программе.

Ключевые слова: информационная безопасность, мониторинг сетевого трафика, спам.

Введение. Средства сетевого мониторинга являются неотъемлемой частью систем администрирования и управления современными информационными комплексами. Они используются при решении ряда задач [1–7] в таких областях, как техническое обслуживание сетей, информационная безопасность, статистические исследования информационных потоков.

Одним из инструментов мониторинга являются датчики, производящие анализ сетевого трафика [3–5]. Их можно разделить на две группы. Узловые датчики отслеживают информационные потоки через компьютеры, образующие сетевые узлы. Линейные датчики контролируют линии связи, соединяющие сетевые узлы, как правило, через зеркальные порты сетевых коммутаторов. Оба типа устройств имеют свои достоинства и недостатки и взаимно дополняют друг друга.

В связи с широким использованием электронной почты как средства коммуникации и сильной «засорённостью» почтового трафика паразитными рекламными сообщениями, называемыми спамом (по данным фирмы "Symantec Corp."(США) до 90 % электронных сообщений являются рекламой [7]), важное место в системах изучения и противодействия этому явлению занимают датчики мониторинга электронной почты (ДМЭП).

Существует два основных подхода к идентификации спама — по формальным признакам сообщения и его содержанию [3]. Формальные методы включают фильтрацию по спискам (почтовым адресам, IP-адресам) или разнообразным параметрам письма (большое количество отправителей, отсутствие получателя, путь, формат, размер и т. д.). Распознавание по содержанию предполагает семантический анализ (контент-анализ текста, выявление специфических словосочетаний, эвристики, анализ по сигнатурам и т. п.), статистические методы (такие как байесовский статистический анализ) или другие стоящие особняком подходы (например, технология Recurrent Pattern Detection фирмы "Commtouch"(США)).

Современные устройства защиты корпоративной электронной почты [8–11] сочетают оба подхода. При этом первоначально производится фильтрация по формальным признакам с целью отсеять подавляющее большинство ненужных сообщений. В дальнейшем производится контекстный отбор. Необходимые признаки фильтрации принимаются

устройством дистанционно из глобальной сети мониторинга. Примерами могут служить: Symantec Global Intelligence Network, собирающая в Интернете информацию о негативной активности с помощью датчиков более чем в 200 странах мира; сеть датчиков компании "Commtouch", осуществляющая глобальный мониторинг сообщений в рамках системы Zero-Hour Virus Outbreak Protection; системы мониторинга компаний CISCO, "McAfee", "Barracuda Networks"(США). Конструктивно устройства представляют собой сложные аппаратно-программные комплексы, включающие компьютеры в стоечном исполнении, оборудованные одним или несколькими ЦПУ, жёсткими дисками, сетевыми адаптерами и работающие под управлением как стандартных, так и специально созданных операционных систем (ОС). Прикладное программное обеспечение выполняет функции анализа и фильтрации входных потоков данных. Управление осуществляется через web-интерфейс и консоль оператора.

Процесс мониторинга электронной почты в современных корпоративных сетях (см., например, [6]) состоит из следующих этапов:

1) считывание сетевых IP-пакетов;
2) составление почтовых сессий из полученных пакетов;
3) выделение почтовых сообщений из сессий;
4) декодирование и анализ почтовых сообщений;
5) архивирование почтовых сообщений.

На практике в задачах сетевого мониторинга популярны решения с применением библиотеки LibPcap в ОС Unix или её варианта для платформы Windows — WinPcap [12, 13]. Данная библиотека предоставляет широкий набор средств чтения пакетов с линии и их первоначальной фильтрации, что значительно облегчает реализацию двух первых этапов мониторинга электронной почты. Создание более интеллектуальных [14] датчиков путём увеличения количества выполняемых этапов позволяет перейти к распределённому мониторингу почтового трафика и за счёт этого повысить суммарную производительность и пропускную способность системы.

Разработка подобных приборов на универсальном оборудовании с использованием широко распространённых операционных систем имеет ряд преимуществ (гибкость, возможность использовать готовые разработки, стоимость комплектующих и т. д.). Однако эти устройства, как правило, уступают специализированным в производительности. Поэтому представляется актуальным рассмотрение различных сторон процесса программной эмуляции линейного интеллектуального ДМЭП на универсальной аппаратно-программной платформе.

В процессе моделирования устройства, обеспечивающего первые четыре этапа процесса мониторинга и выделяющего почтовые сообщения (легальные или спам) по формальным признакам, загружаемым извне [15], исследовались возможности наиболее распространённой в настоящее время платформы Win32 семейства ОС Windows. В рамках исследования, проводимого в среде ОС Windows XP, использовался многопоточный механизм организации обработки данных. Выявлялись особенности и ограничения платформы как основы для создания такого типа устройств. Проведён сравнительный анализ стандартного в ОС Windows инструмента чтения сетевых пакетов — Raw Sockets и библиотеки WinPcap, предоставляющей аналогичные возможности.

Во время экспериментов были определены временны́е характеристики эмулятора, свидетельствующие о его работоспособности в условиях предельной нагрузки на линии связи 100 Мбит/с.

Модель линейного интеллектуального датчика электронной почты. Структурно-функциональная модель интеллектуального сетевого датчика мониторинга построена на основе требований к устройству, вытекающих из его практического применения, и состоит из модуля считывания информации (МСИ), анализатора (А), модуля управления

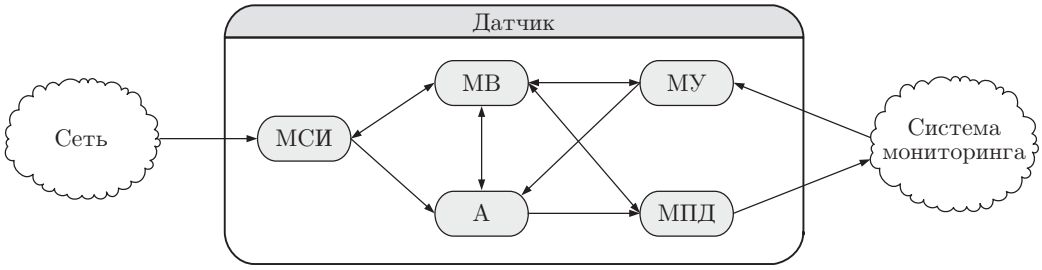

Рис. 1. Структурно-функциональная схема линейного интеллектуального ДМЭП

(МУ), модуля передачи данных (МПД) и модуля визуализации (МВ) (рис. 1). МСИ читает пакеты, циркулирующие в контролируемом сегменте сети, проверяет их целостность и фильтрует по основным почтовым протоколам SMTP, POP3 и IMAP4. Анализатор производит обработку собранной информации и выделяет почтовые сообщения, которые фильтруются по различным критериям отбора: IP-адресу компьютера, электронному адресу отправителя или получателя и т. д. Алгоритм работы анализатора приведён на рис. 2. МПД передаёт устройству управления датчиком (УУД) отобранные электронные письма. МУ организует канал дистанционного управления датчиком со стороны УУД. На основе представленной схемы была создана программная модель-эмулятор для платформы Win32, состоящая из пяти параллельно работающих потоков (по количеству основных модулей).

Для проведения сравнительного анализа возможностей Windows по чтению пакетов с линии связи МСИ был реализован двумя способами: с помощью Raw Sockets и с использованием библиотеки WinPcap.

Особенности определения временны́х характеристик модели. Исследуемую модель можно рассматривать как систему массового обслуживания, входными заявками которой будут пакеты на мониторируемой линии. Важными характеристиками системы массового обслуживания являются время выполнения заявок и коэффициент загрузки. Основа исследуемой модели — Windows XP — является операционной системой с вытесняющей многозадачностью. Это означает, что в любой момент времени ОС может прервать выполнение активного потока программной модели ДМЭП и передать процессор в распоряжение другой задачи. Кроме того, потоки модели ожидают своей очереди для доступа к разделяемым ресурсам, например к общим буферам памяти. Длительности связанных с этим задержек являются непредсказуемыми, трудноизмеримыми и вносят дополнительные проблемы в процесс исследования.

Эксперименты показали, что разброс результатов измерений в последовательности одинаковых испытаний доходил до 5 %. Поэтому использование каких-либо абсолютных временны́х параметров, в том числе и времени выполнения заявок, затруднено.

В связи с изложенным выше для оценки работы программы при различных интенсивностях почтового трафика использовался коэффициент относительной занятости модели процессом мониторинга. Этот параметр, в определённой степени заменяющий коэффициент загрузки, задаётся как отношение измеренного времени обработки программой поступающих сетевых пакетов к временно́му интервалу между обнаружением первого и последнего пакетов. Чем ниже эта величина, тем лучше модель справляется с заданной нагрузкой.

В исследуемой модели имеется три основных модуля, обеспечивающие процесс мониторинга, — это МСИ, анализатор и МПД. Остальные являются вспомогательными и не вносят ощутимого вклада в общее время работы программы. Поэтому коэффициент

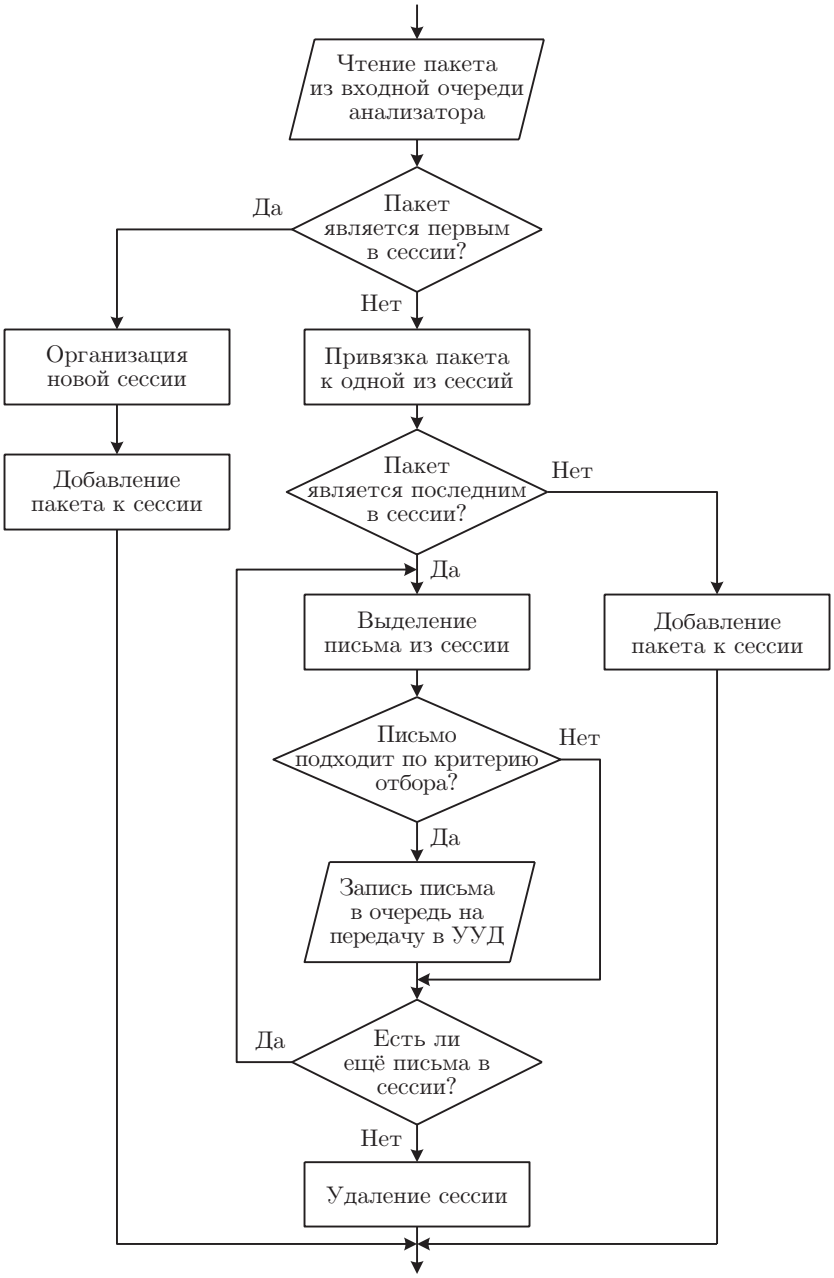

Рис. 2. Блок-схема алгоритма работы анализатора

относительной занятости модели ($L_{\text{м}}$) можно вычислить по формуле

$$L_{\text{м}} = \frac{T_{\text{МСИ}} + T_{\text{А}} + T_{\text{МПД}}}{T_{\text{л}}} \cdot 100, \tag{1}$$

где $T_{\text{МСИ}}$, $T_{\text{А}}$, $T_{\text{МПД}}$ — измеренные времена, потраченные соответствующими потоками на обработку входных пакетов; $T_{\text{л}}$ — время прохождения пакетов в линии связи. Коэффициент выражен в процентах. Получаемые по формуле (1) значения $L_{\text{м}}$ могут служить верхней оценкой способности представленной модели ДМЭП без потерь считывать, анализировать и передавать в систему мониторинга проходящий почтовый трафик определённой интенсивности. Как только значение $L_{\text{м}}$ становится 100 %, это означает, что модель достигла верхнего предела своей производительности.

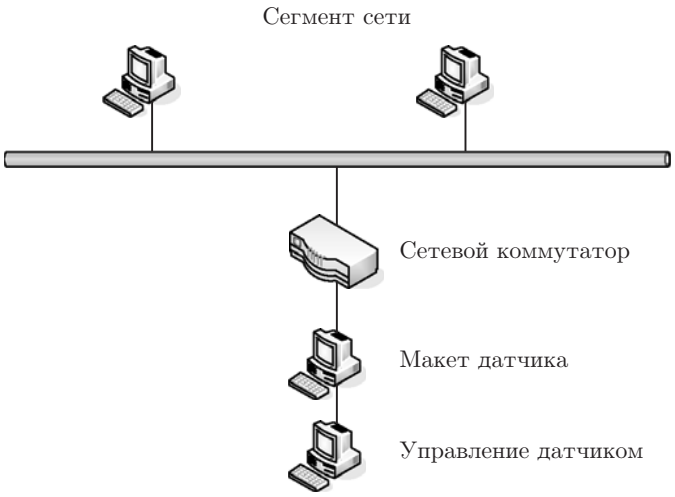

Рис. 3. Схема стенда для испытаний модели линейного интеллектуального ДМЭП

Следует отметить, что представленные числовые значения, полученные в процессе экспериментов с программной моделью датчика, зависят от используемой аппаратной части. Применение другого оборудования может дать другие результаты.

Схема проведения экспериментов по исследованию модели. Испытания модели линейного интеллектуального ДМЭП проводились в рамках тестовой локальной сети 100 Мбит/с (рис. 3).

Программа-эмулятор была размещена на компьютере с двумя сетевыми адаптерами. Адаптер для считывания первичной информации подключён к шлюзу сегмента сети, в котором размещались компьютеры, имитирующие почтовый трафик с заданными характеристиками. Второй адаптер связан с устройством управления датчиком, в качестве которого выступал ещё один компьютер со специализированным программным обеспечением. Компьютеры оснащены процессором Intel Core 2 Duo 2,33 ГГц и памятью 1 Гбайт. В качестве операционной системы использовалась Windows XP. Искусственная нагрузка создавалась путём многократной посылки почтовых сообщений размером 1, 5, 10, 20 или 50 Кбайт. Это продиктовано результатами исследований компании IBM, свидетельствующими о том, что за период с 2005 по середину 2008 г. размер электронного спам-сообщения варьировался от 3 до 11 Кбайт [16]. Процедура выделения рекламных писем выполнялась по IP-адресам отправителя и получателя сообщения.

Проводились следующие испытания в целях определения:

1) времени обработки МСИ пакетов искусственно создаваемого трафика при использовании как механизма Raw Sockets, так и библиотеки WinPcap;

2) характера зависимости коэффициента относительной занятости модели $L_{\text{м}}$ от интенсивности трафика в линии связи $I_{\text{т}}$ при постоянном размере почтового сообщения;

3) вида зависимости коэффициента относительной занятости модели $L_{\text{м}}$ от размера электронного письма при постоянной интенсивности почтового трафика;

4) предельных интенсивностей почтового трафика для почтовых сообщений разных размеров.

Анализ полученных результатов. В процессе испытаний многопоточной модели-эмулятора линейного интеллектуального ДМЭП были получены следующие результаты.

1. Поток МСИ при условии более высокого приоритета по сравнению с остальными потоками модели успевал обработать все пакеты искусственно создаваемого трафика как при использовании Raw Sockets, так и библиотеки WinPcap. Результаты экспериментов сведены в табл. 1. Введём обозначения: величина $\text{Р}_{\text{п}}$ — размер почтового сообщения; $T_{\text{лS}}$

Таблица 1

$Р_п$, Кбайт	Raw Sockets		WinPcap	
	$T_{лS}$, мс	$T_{МСИ S}$, мс	$T_{лW}$, мс	$T_{МСИ W}$, мс
1	11875	11875	11860	11875
5	26250	26250	26219	26219
10	44672	44672	45031	45047
20	73125	73125	73766	73797
50	160969	160969	161078	161094

и $T_{лW}$ — время передачи пакетов по линии связи для библиотек Raw Sockets и WinPcap соответственно; $T_{МСИ S}$ и $T_{МСИ W}$ — интервалы времени от момента чтения первого пакета до момента передачи последнего пакета анализатору для Raw Sockets и WinPcap. В каждом эксперименте передавалось по 30000 почтовых сообщений. Из таблицы видно, что максимальное отклонение временных интервалов МСИ от времени прохождения пакетов составляет 0,126 %.

2. Коэффициент относительной занятости модели $L_м$ прямо пропорционален интенсивности потока в линии связи $I_т$ при постоянном размере почтового сообщения:

$$L_м = K_п I_т, \tag{2}$$

где $K_п$ — коэффициент пропорциональности, зависящий от размера почтового сообщения. Графики зависимостей коэффициента $L_м$ от интенсивности почтового трафика при размерах письма 1, 5, 10, 20 и 50 Кбайт показаны на рис. 4.

3. Коэффициент относительной занятости модели падает с увеличением размера электронного письма при постоянной интенсивности почтового трафика. Это связано с алгоритмическими особенностями работы модели. График зависимости коэффициента $L_м$ от размера почтового сообщения при $I_т = 5{,}5$ Мбайт/с представлен на рис. 5.

4. На имеющемся оборудовании в тестовой локальной сети на линии Ethernet удалось получить максимальную $I_т$, равную 84,8 Мбит/с, при передаче писем размером 50 Кбайт. Значение $L_м$ равнялось 26,6 %. Однако наибольшую нагрузку на датчик дают сообщения размером 1 Кбайт. Для них предельная $I_т$ составила 60,64 Мбит/с, а значение $L_м$

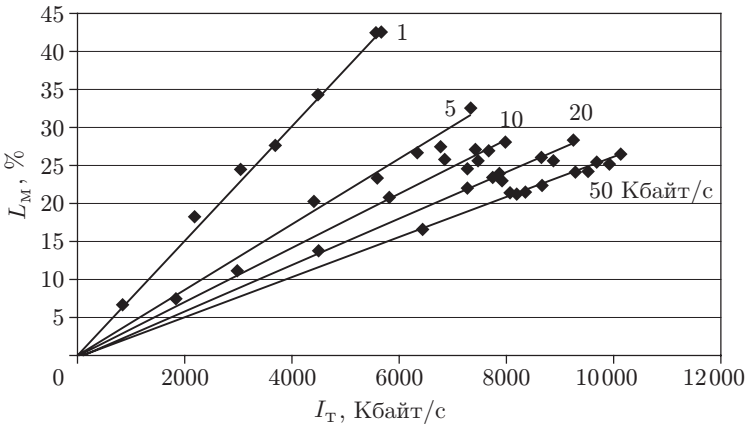

Рис. 4. Зависимости коэффициента относительной занятости модели от интенсивности почтового потока при передаче сообщений

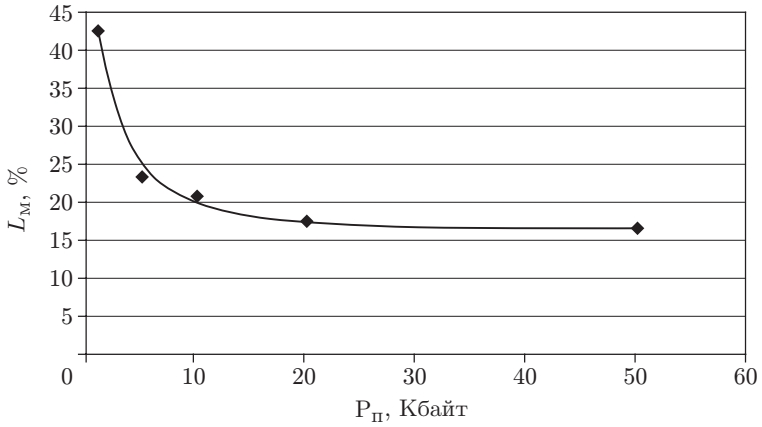

Рис. 5. Зависимость коэффициента относительной занятости модели от размера почтового сообщения при постоянной интенсивности потока

Таблица 2

$Р_п$, Кбайт	$I_т$, Мбит/с	$N_с$, ед./с
1	142,4	5924
5	204,0	3497
10	246,4	2384
20	276,8	1444
50	319,2	701

достигло 42,6 %. Если производить оценку предельной интенсивности трафика для модели, при которой значение $L_м$ будет равняться 100 % для размера сообщения в 1 Кбайт, то по формуле (2) она составит 142,4 Мбит/с. Другая величина, характеризующая работу датчика, созданного на основе модели, — число почтовых сессий, обрабатываемых в секунду. В табл. 2 представлены $I_т$ и количество обрабатываемых в секунду почтовых сессий $N_с$ для сообщений размерами 1, 5, 10, 20 и 50 Кбайт при значении $L_м$, равном 100 % во всех случаях. Из таблицы видно, что предельные для модели данного ДМЭП интенсивности почтового трафика имеют значения, превышающие теоретический предел пропускной способности линии связи 100 Мбит/с.

Заключение. Испытания модели-эмулятора линейного интеллектуального датчика мониторинга электронной почты на платформе Win32 при применении многопоточного механизма с алгоритмом фильтрации сообщений по формальным признакам показали её работоспособность. Операционная система Windows XP при использовании современной аппаратной части может реализовываться как основа для создания ДМЭП, рассчитанного на линии связи 100 Мбит/с. Для более скоростных линий требуются более мощные процессоры или аппаратная реализация некоторых компонентов модели, например модуля считывания информации. Результаты сравнения библиотек WinPcap и Raw Sockets показывают, что характеристики датчика аналогичны. Использование WinPcap в программном обеспечении устройства облегчает замену платформы Windows платформой Unix в случае необходимости. ДМЭП, созданный на основе данной модели, может использоваться как составная часть систем информационной безопасности для борьбы со спамом или аудита электронной почты, а также в системах статистических исследований информационных потоков.

СПИСОК ЛИТЕРАТУРЫ

1. **Уилсон Э.** Мониторинг и анализ сетей. Методы выявления неисправностей. М.: «Лори», 2002. 364 с.
2. **Жижченко А. Б., Васенин В. А.** Алгоритмическое и программное обеспечение интернета следующего поколения // Информационное общество. 2005. Вып. 1. С. 56–64.
3. **Запечников С. В., Милославская Н. Г., Толстой А. И., Ушаков Д. В.** Информационная безопасность открытых систем. Т. 2. Средства защиты в сетях. М.: Горячая линия—Телеком, 2008. 558 с.
4. **Сердюк В. А.** Сбор данных системами обнаружения атак // BYTE/Россия. 2003. № 2. С. 74–78.
5. **Ушаков Д. В.** Функциональные возможности современных систем обнаружения вторжений // Безопасность информационных технологий. 2005. № 1. С. 24–31.
6. **Слепов О.** Спам: мониторинг электронной почты // Открытые системы. 2004. № 10. С. 36–40.
7. **The state of spam.** A Monthly Report // Symantec corp. Report #30, June 2009. 13 p. URL: http://eval.symantec.com/mktginfo/enterprise/other_resources/b-state_of_spam_report_06-2009.en-us.pdf (дата обращения: 02.09.2010).
8. **IronPort** C650 Email security appliance for large enterprises and ISPs. Datasheet // CISCO SYSTEMS. 2007. 5 p. URL: http://www.ironport.com/pdf/ironport_c650_datasheet.pdf (дата обращения: 02.09.2010).
9. **Zero** hour virus outbreak protection a key layer in complete enterprise email security // Commtouch Software Ltd. 2007. 9 p. URL: http://www.commtouch.com/downloads/Commtouch_Zero-hour_in_Gateway_WP.pdf (дата обращения: 02.09.2010).
10. **McAfee** email and web security appliance. Datasheet // McAfee, Inc. 2009. 2 p. URL: http://mcafee.com/us/local_content/datasheets/ds_email_web_security_appliance.pdf (дата обращения: 02.09.2010).
11. **Barracuda** spam & virus firewall. Datasheet // Barracuda Networks Inc. 2010. 2 p. URL: http://www.barracudanetworks.com/ns/downloads/Datasheets/Barracuda_Spam_&_Virus_Firewall_DS_US.pdf (дата обращения: 02.09.2010).
12. **WinPcap-based** tools and programs. URL: http://www.winpcap.org/misc/links.htm#tools (дата обращения: 02.09.2010).
13. **Risso F., Degioanni L.** An architecture for high performance network analysis // Proc. of the 6th IEEE Symposium on Computers and Communications (ISCC 2001). Hammamet, Tunisia. July 2001. P. 686–693.
14. **Mintchell G. A.** Use sensors intelligently // Control Eng. 2001. **48**, N 10. P. 39–42.
15. **Будников К. И., Клисторин И. Ф., Курочкин А. В., Лылов С. А.** Датчик удалённого мониторинга электронной почты // Датчики и системы. 2008. № 9. С. 35–37.
16. **IBM Internet** security systems X-force®2008 mid-year trend statistics // IBM Global Technology Services. 2008. 81 p. URL: http://www-935.ibm.com/services/us/iss/xforce/trendreports/xforce-2008-annual-report.pdf (дата обращения: 02.09.2010).

Поступило в редакцию 28 апреля 2010 г.