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

Email: budnikov@iae.nsk.su

МЕТОД ЭКСПЕРИМЕНТАЛЬНОЙ ОЦЕНКИ ДАТЧИКОВ МОНИТОРИНГА ЭЛЕКТРОННОЙ ПОЧТЫ

В статье рассмотрен метод экспериментальной оценки датчиков мониторинга электронной почты, заключающийся в определении характеристик приборов при обработке ими искусственной нагрузки, которая учитывает статистические свойства реальных потоков почтовых данных, циркулирующих в сети Интернет. Предложены 3 типа нагрузки, характерные для разных способов включения датчика в сеть. Рассмотрена методика генерации нагрузочного трафика.

Ключевые слова: информационная безопасность, мониторинг сетевого трафика, спам, экспериментальная оценка производительности.

Введение

Электронная почта занимает одно из центральных мест среди средств электронных коммуникаций, применяемых как организациями, так и частными лицами. В 2010 г. в мире насчитывалось около 3 млрд активных почтовых адресов. Ожидается, что к 2014 г. их количество будет приближаться к 4 млрд [1]. Подобная популярность электронной почты способствовала распространению такого явления, как спам (массовая рассылка бесполезных рекламных сообщений), наносящий вред как получателям, так и интернет-провайдерам, уменьшая эффективность использования их временных и аппаратных ресурсов. Согласно исследованиям Symantec Corp. (США) до 90 % электронных сообщений являются рекламой [2]. Современные устройства защиты корпоративной электронной почты [3] используют два основных подхода для борьбы со спамом: фильтрацию по формальным признакам почтового сообщения (например, по почтовым адресам, IP-адресам) и по его содержанию. Необходимые признаки фильтрации загружаются в устройства дистанционно из глобальной сети мониторинга, которая представляет собой распределенную сеть датчиков, собирающих информацию о негативной активности в Интернете, и ряда аналитических центров, выполняющих функции управления и координации. Примерами могут служить системы американских компаний Symantec, CISCO, McAfee, Commtouch, Barracuda Networks. Датчик мониторинга

[1] Email Statistics Report, 2010 // The Radicati Group, inc., URL: www.radicati.com (дата обращения 02.09.2011).

[2] State of Spam & Phishing. A Monthly Report // Symantec corp. Report #53, May 2011, 12p. http://www.symantec.com/content/en/us/enterprise/other_resources/b-state_of_spam_and_phishing_report_05-2011.en-us.pdf (дата обращения 02.09.2011).

[3] 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; 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; 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; 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 02.09.2011).

электронной почты (ДМЭП) в подобных системах представляет собой программно-аппаратный комплекс, обеспечивающий разбор входного сетевого трафика, выделение и анализ почтовых сессий в соответствии с заданными критериями и алгоритмами, передачу результатов анализа на следующие уровни системы мониторинга.

Важнейшими параметрами ДМЭП являются: наибольшая интенсивность контролируемого трафика, при которой датчик обрабатывает электронную корреспонденцию без потерь; коэффициент загрузки устройства; временная задержка, с которой датчик передает собранную информацию в систему мониторинга, и др. Эти величины могут быть измерены экспериментально в процессе обработки искусственной нагрузки с заданными характеристиками, которая создается генератором на входе датчика. В рамках исследования модели ДМЭП на платформе Win32 [1–3] выяснилось, что при одинаковой интенсивности почтового трафика коэффициент загрузки устройства зависит от размера почтового сообщения. В связи с этим для правильного измерения параметров датчика большое значение имеют состав и статистические пропорции тестовой нагрузки, которые должны отражать реальные условия эксплуатации прибора.

Среди множества способов генерации Интернет-трафика [4] можно выделить подходы одной из наиболее известных организаций, работающих в области экспериментальной оценки производительности вычислительных систем – Standard Performance Evaluation Corporation (SPEC). Для генерации почтовых потоков ею разработаны широко известные тесты SPEC MAIL2001 и SPEC MAIL2009, эмулирующие процесс электронной переписки в рамках организации и позволяющие оценить эффективность работы корпоративных почтовых серверов. Характеристики этого трафика соответствуют составу и статистическим свойствам информационных потоков между почтовыми клиентами и сервером. Однако использование данных тестов для определения характеристик ДМЭП не корректно, поскольку в них не учитываются такие свойства сетевого трафика, как наличие спама, доходящего до 90 % и имеющего свои статистические характеристики, а также количественное соотношение почтового трафика с другими видами информационных потоков в Интернете. Поэтому при проведении экспериментальной оценки ДМЭП необходимо учитывать особенности трафика в местах подключения его к линиям связи. Датчик может быть установлен (а) в исследовательской лаборатории для получения его пиковых характеристик при обработке почтовых сообщений разного размера, (б) в непосредственной близости (в одной локальной сети) от почтового сервера, где циркулирует преимущественно почтовый трафик, или (в) может быть подключен к магистральным каналам сети Интернет с различной пропускной способностью, где почтовый трафик – лишь часть общего Интернет-трафика.

Поэтому в рамках исследования модели ДМЭП, проводимых в Институте автоматики и электрометрии (ИАиЭ) СО РАН, были разработаны методы экспериментальной оценки датчиков [3]. Они отражают особенности сетевого трафика, с которым впоследствии предстоит работать данным устройствам, и представлены в виде теста, состоящего из трех частей.

Статистические закономерности почтового трафика

Необходимые статистические исследования почтового трафика проводились на базе офисной подсети одного из подразделений ИАиЭ СО РАН, имеющей выход в Интернет. Были проанализированы данные, собранные за 2010 г. и охватывающие примерно 18 000 электронных сообщений. Инструментом сбора информации служил ДМЭП, разработанный в ИАиЭ [1].

На рис. 1 представлено полученное распределение электронных писем по размерам. По оси абсцисс отложена длина сообщения, по оси ординат – его процентное содержание в общем количестве. Средний размер письма, определяемый как математическое ожидание, составил 12 411 байт. Более 94 % писем имеют размер, не превышающий 50 кБ. В общем объеме трафика

[4] Traffic Generators for Internet Traffic. URL: http://www.icir.org/models/trafficgenerators.html; Standard Performance Evaluation Corporation. SPEC MAIL2001. URL: http://www.spec.org/mail2001/; Standard Performance Evaluation Corporation. SPEC MAIL2009. URL: http://www.spec.org/mail2009/ (дата обращения 02.09.2011).

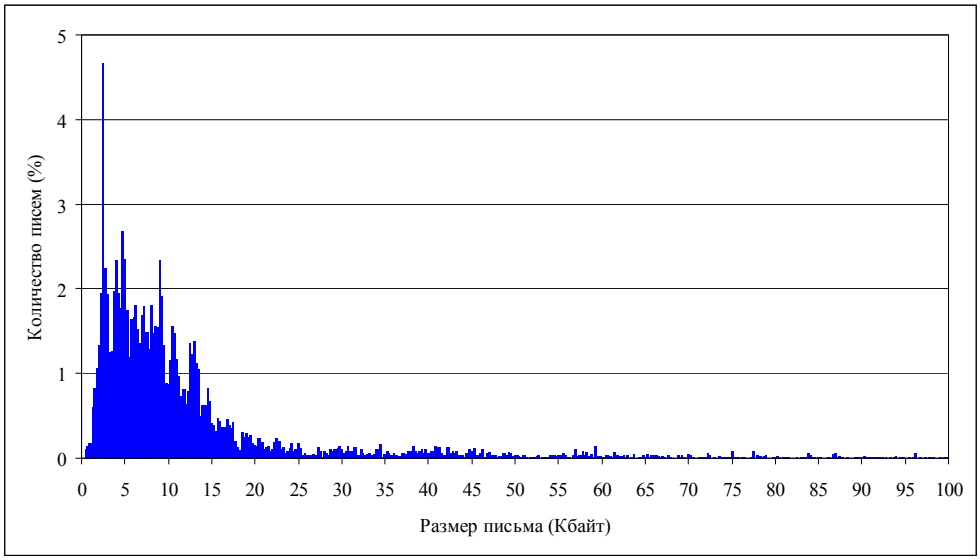

Рис. 1. Распределение электронных писем по их размерам за 2010 г.

в мониторируемой подсети электронная почта заняла порядка 1 %. Поскольку представленные результаты коррелируют с данными зарубежных компаний, исследующих интернет-трафик, например Ipoque GmbH (Германия) [5] и IBM (США) [6], то они были взяты за основу экспериментальной оценки ДМЭП.

Методика экспериментальной оценки датчика

Разработанная методика оценки ДМЭП получила название TransMail (TM) и состоит из 3 частей (TM-A, TM-B и TM-C), которые соответствуют различным способам подключения ДМЭП. Поскольку статистические исследования проводились в 2010 г., то построенный на их основе вариант теста обозначен как TransMail-2010.

TM-A – тест, предназначенный для исследований пиковых характеристик датчика, установленного в исследовательской лаборатории. На вход прибору подается однородный почтовый поток с сообщениями единого размера, например 1, 2, 4, 8, 16, 32 или 64 кБ. Данный тест позволяет определить характер изменений и предельные величины параметров прибора в зависимости от размера почтового сообщения.

TM-B – тест, в котором датчик анализирует почтовый трафик со случайным распределением сообщений по длине, что соответствует размещению прибора в непосредственной близости к почтовому серверу, где поток преимущественно состоит из почтовых сообщений. Необходимое распределение формируется в процессе предварительного статистического исследования реального потока. На основе изложенных выше результатов для офисной подсети при генерации тестового трафика можно использовать распределение, представленное на рис. 2. Оно имеет логарифмическую шкалу размеров сообщений, которая более удобна для данной задачи, чем линейная (см. рис. 1).

Процентное содержание почтовых сообщений разных размеров в тестовой смеси представлено в табл. 1.

TM-C – тест, в котором датчик анализирует поток, эмулирующий трафик, характерный для сети Интернет. В нем почтовая составляющая имеет только 1 % от общего трафика. Состав почтовой части трафика такой же, как и в тесте TM-B (см. табл. 1). Состав не почтовой части трафика не регламентируется.

[5] *Schulze H., Mochalski K.* Internet Study 2007. The Impact of P2P, File Sharing, Voice over IP, Skype, Joost, Instant Messaging, One-Click Hosting and Media Streaming such as YouTube on the Internet // http://www. ipoque.com/sites/default/files/mediafiles/documents/internet_study_2007.pdf (дата обращения 02.09.2011).

[6] IBM X-Force® 2010 Trend and Risk Report // http://ibm.sharedvue.net/sharedvue/assets/594/ibm_showcase/services/en/113/ (дата обращения 02.09.2011).

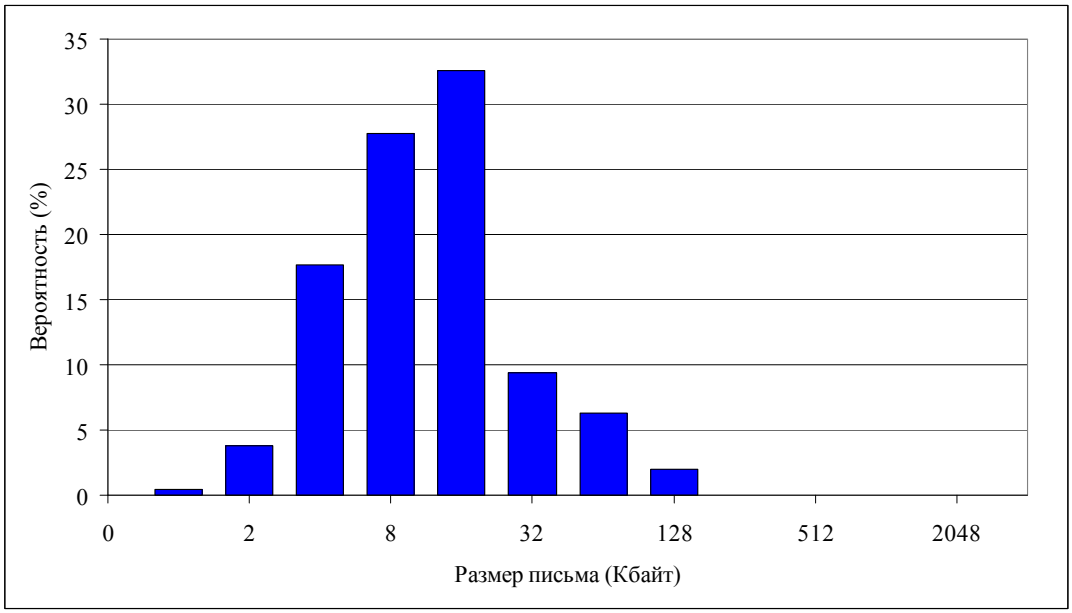

Рис. 2. Распределение вероятности появления электронных писем в тестовом потоке по их размерам

Таблица 1

Процентное содержание электронных писем разных размеров в тестовом потоке

Размер письма (кБ)	1	2	4	8	16	32	64	128
Содержание в тестовой смеси (%)	0,43	3,82	18,32	26,88	32,68	9,71	6,31	1,85

Схема стенда для проведения испытания датчика

Экспериментальная оценка ДМЭП производится в рамках локальной сети, изображенной на рис. 3. К коммутатору, обеспечивающему зеркалирование портов, подключаются ДМЭП и два компьютера, эмулирующих работу почтового сервера и клиента. При этом датчик должен получать все сетевые пакеты, которыми обмениваются данные компьютеры. Управление работой ДМЭП осуществляется со специального устройства управления датчиком (УУД), в качестве которого выступает дополнительный компьютер с соответствующим ПО. Сервер и клиент создают тестовый трафик (почтовый и при необходимости не почтовый), поступающий на вход датчика. ДМЭП выделяет почтовые сообщения из данного потока и передает их в УУД.

Метод генерации почтового трафика

В основу метода генерации тестового почтового трафика положен принцип многократного повтора предварительно записанного реального диалога между почтовым сервером и клиентом. На подготовительном этапе при помощи специализированного ПО формируются файлы с почтовыми сессиями, необходимыми для планируемых испытаний. Каждый файл содержит последовательность сообщений одной сессии с признаками направления передачи – от сервера к клиенту или наоборот. Эти файлы размещаются на серверном и клиентском компь-

ютерах, и далее соответствующие программы на их основе создают тестовую нагрузку на входе датчика. Например, для проведения рассмотренных выше испытаний создан набор файлов с сессиями основных почтовых протоколов POP3, SMTP и IMAP, в которых пересылаются письма размером от 1 до 150 кБ.

Алгоритм генерации почтовой сессии со стороны клиента представлен на рис. 4.

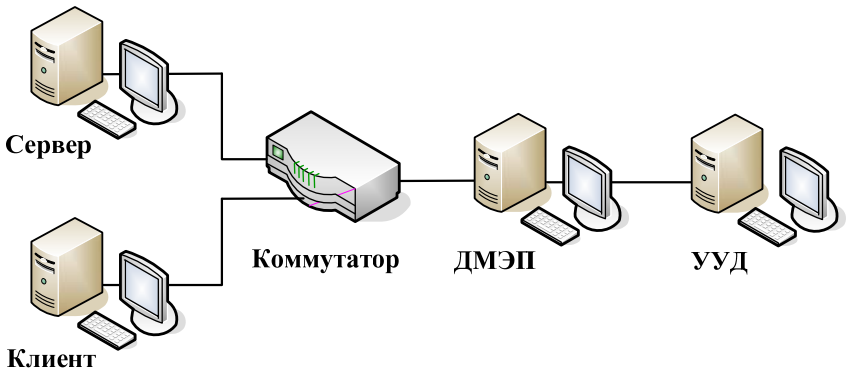

Рис. 3. Схема стенда для проведения испытаний ДМЭП

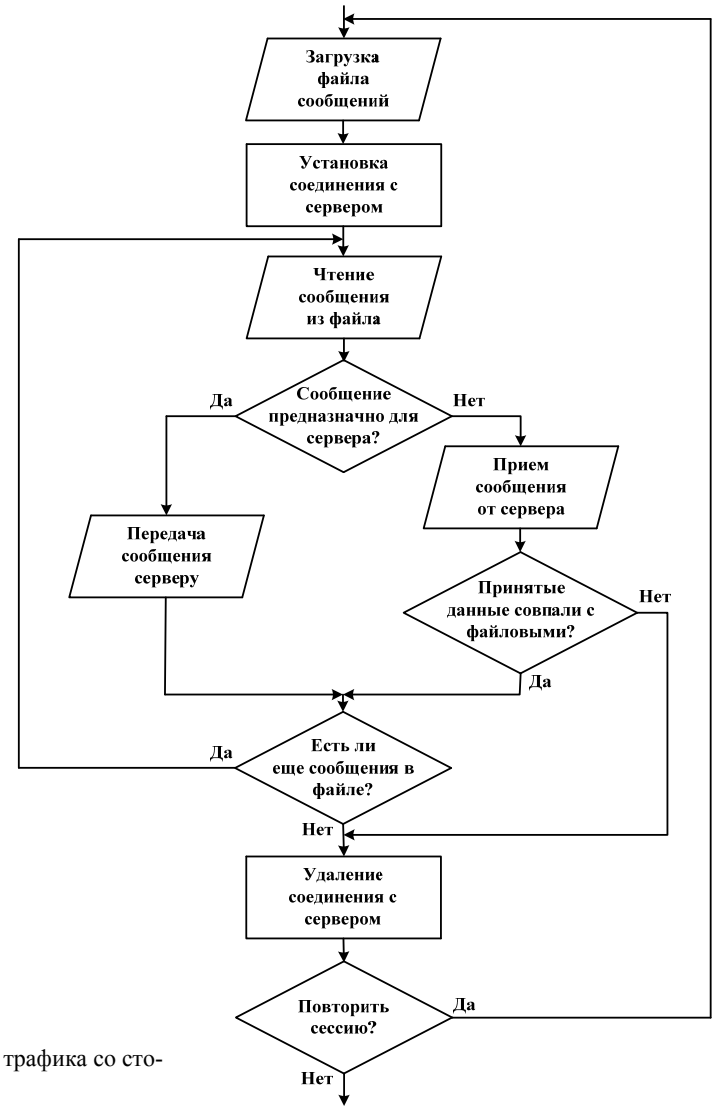

Рис. 4. Алгоритм генерации почтового трафика со стороны клиента

Процесс начинается с загрузки соответствующего файла сообщений. Далее устанавливается соединение с сервером. Из файла считывается первое сообщение и проверяется направление передачи. Если оно должно быть передано серверу, то программа отправляет данные. В противном случае клиент принимает сообщение от сервера и сравнивает его с информацией из файла. При обнаружении различий диалог прекращается. Аналогичным образом обрабатываются все последующие сообщения сессии. После этого соединение с сервером разрывается, и при необходимости инициируется новая почтовая сессия. Примерно по такой же схеме работает и серверное приложение. Однако между ними имеются два отличия. Во-первых, программы различаются механизмом установления соединения. Во-вторых, в эмуляторе клиента, могут создаваться десятки независимых потоков (thread в концепции Windows API), работающих по представленному алгоритму, тогда как со стороны сервера – только один для каждого поддерживаемого почтового протокола.

Рассмотренный способ позволяет обеспечить почтовый трафик для проведения любых тестов из предложенного метода сравнительной оценки. С его помощью можно создавать и «паразитную» составляющую тестовой нагрузки. Для этого необходим дополнительный сервер, привязанный к непочтовому TCP-порту, а клиентские потоки направляют определенную часть своих обращений на данный сервер.

Интенсивность тестового трафика в представленном методе регулируется количеством потоков в клиентской программе. На рис. 5 изображены графики этой зависимости для однородных потоков на основе писем с размерами 1, 2, 4, 8 и 32 кБ. Измерения проводились в локальной сети 1 ГБит/с на базе коммутатора ProCurve 2510G. В качестве сервера и клиента использовались компьютеры с материнской платой Intel на базе чипсета Intel P7P55, 4-ядерным процессором Intel Core i5 2,67 ГГц, сетевой платой Intel 1000 GT, оперативной памятью 4 ГБ и операционной системой Windows XP.

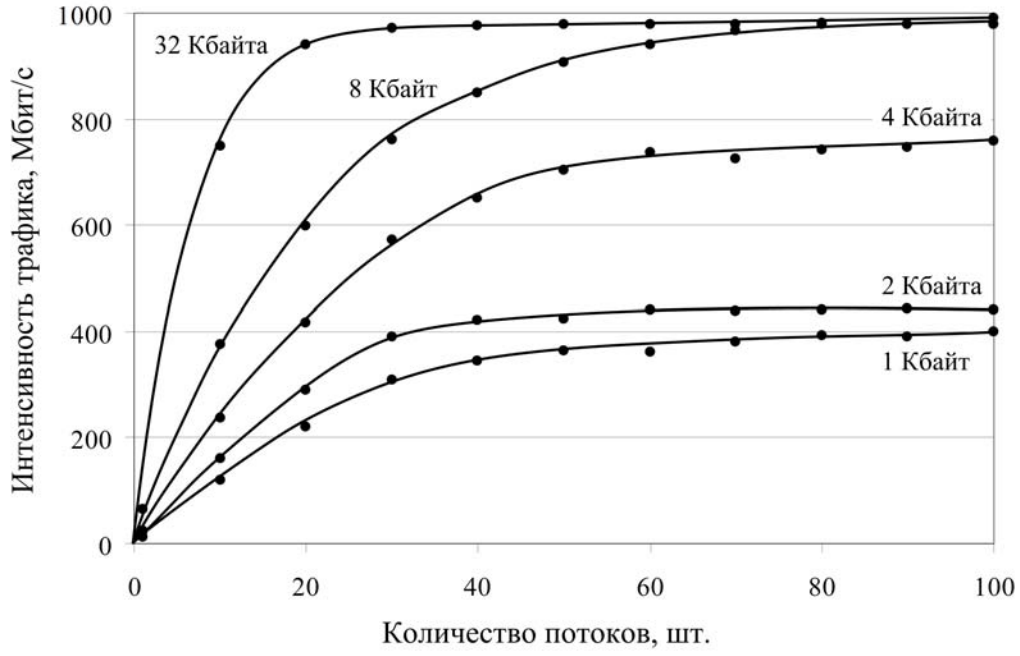

Рис. 5. Зависимость интенсивности тестового трафика от количества потоков в клиентской программе при передаче писем с размером 1, 2, 4, 8 и 32 КБ

Заключение

Для экспериментальной оценки датчиков мониторинга электронной почты предложен метод определения параметров этих приборов при обработке ими искусственной нагрузки, эмулирующей реальные потоки в сети Интернет, с которыми устройствам приходится иметь дело в процессе эксплуатации. Структура нагрузки основана на статистических характеристиках реальных потоков. При помощи данного метода моделируются три различных способа подключения прибора к линии связи: (а) в исследовательской лаборатории для получения его пиковых характеристик при обработке почтовых сообщений разного размера, (б) в непосредственной близости (в одной локальной сети) от почтового сервера, где циркулирует преимущественно почтовый трафик, или (в) может быть подключен к магистральным каналам сети Интернет различной пропускной способности, где почтовый трафик – лишь часть общего Интернет-трафика.

Унификация нагрузки позволяет оценить и сравнить устройства разных производителей и, таким образом, определить наиболее подходящий прибор для выполнения требуемой задачи.

Предложенная методика генерации нагрузочного трафика позволяет создать трафик для всех изложенных видов тестов, а также для статистических смесей с иными пропорциями. Ее недостатком является дискретность изменения величины интенсивности генерируемого трафика, которая регулируется количеством эмулирующих потоков между клиентом и сервером. Однако данное обстоятельство не мешает экспериментально провести необходимые сравнительные процедуры.

Список литературы

1. *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.

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

3. *Будников К. И., Клисторин И. Ф., Курочкин А. В.* Исследование многопоточной модели линейного интеллектуального датчика мониторинга электронной почты на платформе Win32 // Автометрия. 2010. № 10. С. 124–131.

Материал поступил в редколлегию 16.11.2011

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

A METHOD FOR EXPERIMENTAL ESTIMATION OF E-MAIL MONITORING SENSORS

The paper describes a method for experimental estimation of email monitoring sensors. Characteristics of a device are obtained while it processes an artificial workload, which parameters reflect statistical properties of real Internet email data streams. Three types of workload are suggested that correspond to 3 different types of a sensor connection to a communication network. A method of workload traffic generation is considered.

Keywords: information security, network traffic monitoring, spam, experimental estimation of performance.