ИЗВЕСТИЯ ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЙ
Т. 56, № 9/2 ФИЗИКА 2013

УДК 519.2; 004.7; 004.67; 004.492.3

Н.Г. БУЛАХОВ, А.А. СПИРИДОНОВ

РАСЧЁТ И ПРЕДСТАВЛЕНИЕ ЭНТРОПИЙНЫХ ХАРАКТЕРИСТИК СЕТЕВЫХ ПОТОКОВ В МАСШТАБИРУЕМЫХ КОМПЬЮТЕРНЫХ СЕТЯХ

Предложены способ и реализация представления энтропийных характеристик сетевых потоков, пригодных для сравнения этих характеристик в сетях различного масштаба.

Ключевые слова: *телекоммуникационные системы, информационные потоки, энтропия.*

С каждым годом возрастает роль компьютерных сетей в современном обществе, совершенствуются голосовые и видеокоммуникации, всё больше задействуются электронные средства оплаты, набирает обороты сетевая коммерция. Организации и частные лица становятся зависимыми от распределённых удалённых электронных сервисов. Услуги телекоммуникаций переходят от вспомогательных в разряд первостепенных по важности. Именно поэтому с каждым годом всё больше внимания уделяется вопросам сетевой безопасности, администрирования сетей и анализу передаваемых сетевых потоков [1]. Возникает множество прикладных задач по обеспечению соблюдения корпоративных стандартов предприятий при использовании служебных телекоммуникационных ресурсов, выявлению злоумышленников, использующих вредоносное сетевое программное обеспечение, сбору статистики использования сети и тому подобных. Большинство из этих задач можно решить при использовании энтропийного анализа, характеристик информационных потоков в компьютерных сетях.

Современные компьютерные сети представляют собой унифицированную среду передачи гетерогенных информационных потоков множества пользователей, использующих различное программное обеспечение и широкий спектр протоколов прикладного уровня. Для обеспечения безопасности передаваемая информация часто шифруется. Вместе с этим в качестве транспортного агента самое широкое распространение получило семейство протоколов TCP/IP благодаря своей универсальности, децентрализованности, надёжности и относительной простоте. В таких сетях информационные потоки представляют собой очередь пакетов данных, отправляемых от отправителя к получателю, снабжённых всей необходимой маршрутной и технической информацией.

С точки зрения статистического анализа, каждый отдельный пересылаемый по сети информационный пакет имеет ряд характеристик в виде полей заголовков этого пакета [1]. Конкретное значение поля заголовка информационного пакета можно рассматривать как некоторый символ в последовательности или конкретную реализацию случайного события. Согласно Кл. Шеннону, энтропия определяется для случайного события x с n дискретными возможными реализациями $1,\ldots n$, ассоциированными с дискретным распределением вероятностей. На практике можно ассоциировать символ выходной последовательности потока с реализацией рассматриваемого события и рассматривать случайное событие как процесс, который генерирует один символ или их последовательность:

$$H(x) = -\sum_{i=1}^{n} p(i) \log_2 p(i) . \qquad (1)$$

Использование в данном случае величины информационной энтропии IP- и MAC-адресов, а также TCP- и UDP-портов источника и назначения позволяет установить, например, факт наличия в информационном потоке данных, генерируемых реплицирующимися программами, а также определить источник этих данных [2, 3].

В процессе работы авторами была разработана и реализована система разделения потоков данных [4] на основе предложенного энтройпийного метода [3]. Система в режиме реального времени анализирует значения энтропии параметров передаваемых по сети пакетов [5]. Это позволяет рассматривать только заголовки пакетов и игнорировать их содержимое, что дает возможность обойти проблемы, связанные с шифрованием передаваемых данных, а также увеличить скорость обработки данных.

При детальном рассмотрении было выявлено, что можно провести разделение информационных потоков в зависимости от класса приложения, которое является источником пересылаемых в

потоке данных. То есть по энтропийным характеристикам потока данных с высокой точностью (более 95 %) можно предсказать, какие именно программы являются источником данного потока.

На практике при измерении энтропийных характеристик потоков в сетях разных масштабов может наблюдаться разное количество одновременных участников сетевого взаимодействия. От данного параметра будет, в свою очередь, зависеть максимальная величина информационной энтропии потоков, реализуемая в случае равновероятного распределения реализаций параметров информационных пакетов в потоке согласно определению Кл. Шеннона. Данная особенность не была учтена ранее, что сужало спектр идентифицируемых программ. В случае если для детектируемого класса программного обеспечения характерно установление связей с фиксированным количеством участников, можно применять метод энтропийного анализа непосредственно. В случае если программное обеспечение задействует при передаче данных количество компьютеров, пропорциональное масштабам сети, дифференциация программ внутри данного класса непосредственно методом энтропийного анализа затруднена. Для решения проблемы авторами предлагается разделить полученную для потока величину энтропии на максимальное её значения для данного количества участников сетевого взаимодействия.

Нетрудно видеть, что максимально возможное значение энтропии, реализуемое в случае равномерного распределения вероятностей реализации случайного события, будет зависеть только от количества возможных реализаций и вычисляться по формуле

$$H_{\max}(x) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n \qquad (2)$$

В свою очередь, количеством возможных реализаций случайного события в нашем случае будет число различных вариантов некоторого параметра, встреченного в потоке данных среди заданного количества информационных пакетов. Поделив рассчитанное абсолютное значение энтропии на его максимально возможную величину для данной выборки, получим относительную величину h, характеризующую степень случайности наблюдаемого параметра в диапазоне от 0 до 1 независимо от масштабов сети и количества участников информационного обмена:

$$h(x) = \frac{H(x)}{H_{\max}(x)}. \qquad (3)$$

Такая модификация даёт возможность дифференциации программ внутри класса задействующего при передаче данных количество компьютеров, пропорциональное масштабам сети. Тем самым обеспечивается возможность фильтрации несанкционированных сетевых потоков и обеспечения необходимого качества обслуживания высокоприоритетным потокам даже в условиях шифрования сообщений и маскировки данных под другие виды.

Тестовая реализация предложенного способа выполнена под ОС FreeBSD, имеет модульную легко дополняемую архитектуру (рис. 1) и веб-интерфейс. Это позволяет использовать данную программную реализацию непосредственно на программных маршрутизаторах с целью анализа и управления сетевыми потоками данных, легко адаптируя её к конкретным ситуациям и потребностям.

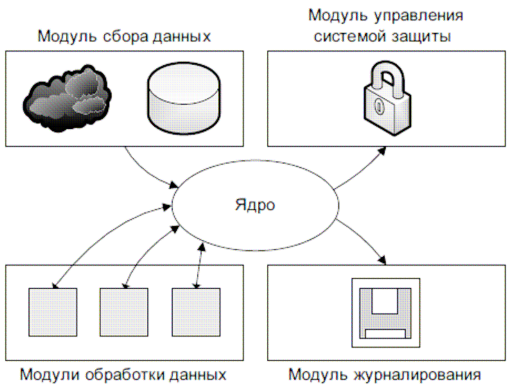
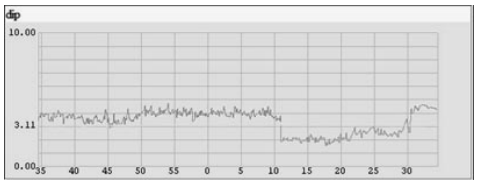


Рис. 1. Модульная структура программной реализации

Визуализация данных производится при помощи графического представления временны**ы**х диаграмм изменения энтропийноых характеристик общего потока данных и табличного представ-

ления наборов текущих энтропийных характеристик потоков от каждого источника данных, определяемого его IP-адресом.



	smac	dmac	sip	dip	sport	dport	packlen	packetflow	Type of traffic
192.168.99.37	0.999999	0.999999	0	0	0.999966	0.999966	0.999999	5659	
192.168.99.33	0	0.614725	0	0.614725	0.735643	0.735643	2.63069	408	
192.168.99.39	0.999998	0.999998	0	0.00693693	1.00645	1.00621	0.161403	8343	

а *б*

Рис. 2. Примеры представления данных: *а* – временная диаграмма изменения энтропийных характеристик общего потока данных; *б* – табличное представление наборов текущих энтропийных характеристик потоков от каждого источника данных, определяемого его IP-адресом

В перспективах доработки рассмотренного метода дифференцирования трафика телекоммуникационных систем дальнейшее исследование энтропийных характеристик различного программного обеспечения, повышение точности определения принадлежности информационных потоков и расширение базы детектируемого программного обеспечения. Также к перспективам относится коммерческая программная реализация данного метода, пригодная к использованию в сетях провайдеров Интернета, предприятий и частных домашних.

СПИСОК ЛИТЕРАТУРЫ

1. Булахов Н.Г., Подгорный Е.А. // Изв. вузов. Физика. – 2008. – Т. 51. – № 9/3. – С. 175–176.
2. Булахов Н.Г. // Изв. вузов. Физика. – 2010. – Т. 53. – № 9/3. – С. 194–195.
3. Булахов Н.Г. // Изв. вузов. Физика. – 2012. – Т. 55. – № 9/2. – С. 319–320.
4. Булахов Н.Г., Спиридонов. А.А., Минаев А.И., Крестьянников А.С. // Изв. вузов. Физика. – 2012. – Т. 55. – № 9/2. – С. 325–326.
5. Булахов Н.Г., Спиридонов. А.А., Минаев А.И., Крестьянников А.С. // Изв. вузов. Физика. – 2012. – Т. 55. – № 9/2. – С. 331–332.

Национальный исследовательский Томский государственный университет,
г. Томск, Россия
E-mail: nik@rff.tsu.ru

Поступила в редакцию 15.07.13.

Булахов Николай Георгиевич, ст. преподаватель;
Спиридонов Александр Алексеевич, студент.

N.G. BULAKHOV, A.A. SPIRIDONOV

CALCULATION AND EXPRESSION OF ENTROPY CHARACTERISTICS OF NETWORK TRAFFIC IN A SCALABLE COMPUTER NETWORKS

A method and realization of the representation of entropy characteristics of network flow suitable for the comparison of these characteristics in the networks of various sizes.

Keywords: *telecommunication systems, information flows, entropy.*