

УДК 519.2, 004.7, 004.67, 004.492.3

Н.Г. БУЛАХОВ, В.Д. МЕДВЕДЕВ, В.С. КУЗНЕЦОВ

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ ЭНТРОПИЙНОГО МЕТОДА АНАЛИЗА СЕТЕВЫХ ПОТОКОВ

Рассматриваются варианты и особенности применения метода энтропийного анализа параметров пакетов сетевых потоков для решения прикладных задач управления компьютерными сетями и описания процессов в них.

Ключевые слова: энтропийный анализ, компьютерные сети, сетевые потоки.

Трудно представить современный мир без компьютерных сетей. К такому широкому распространению данной технологии привели несколько обстоятельств. До компьютерных сетей информацию передавали другими часто узкоспециализированными способами, пригодными для конкретной реализации той или иной технологии, будь то телеграф, телефонная связь, телевидение или военная связь. Прогресс и неизбежные изменения в технологиях приводили к отмиранию одних стандартов и смене их другими, что влекло в свою очередь вынужденную замену оборудования устаревшего морально, но не физически. Кроме того, для обеспечения функционирования нескольких видов связи приходилось строить несколько несовместимых между собой сетей, таких, как сети телевизионных станций и автоматических телефонных станций. Компьютерные же сети изначально задумывались как универсальный способ передачи цифровых данных, поэтому со временем они объединили в себе и поглотили практически все виды предшествующих технологий. Универсальность протокола TCP/IP позволила передавать мультимедийные потоки и отдельные объёмы данных без привязки к конкретной технологии, в то же время индивидуально обеспечивая приемлемое качество обслуживания согласно требованиям программного обеспечения, генерирующего эти данные. Также со временем появились виды человеческой деятельности, связанные исключительно с сетевыми технологиями. Это социальные сети, электронная коммерция, целевая реклама в сети, поиск информации, мобильные службы и так далее. Всё это стало неотъемлемой частью жизни современного общества.

Вместе с тем появились и негативные составляющие сетевого образа жизни, такие, как нецелевое использование рабочего времени, навязчивые массовые рассылки нежелательных электронных сообщений, распространение по сети вредоносного программного обеспечения, скрытое удалённое управление миллионами компьютеров ничего не подозревающих пользователей и тому подобное. Вопросы сетевой безопасности и соблюдения административных политик становятся тем острее, чем критичнее для бизнеса предоставляемый сетевой сервис.

Практика показала, что компьютерные сети эволюционировали в масштабную и сложную систему, что, в свою очередь, усложняет вопросы управления и описания процессов в этих сетях. Усложняет ситуацию и то, что администрирующему субъекту не всегда подчинены и доступны компьютеры, генерирующие сетевые потоки, а передаваемая в них информация может быть зашифрована. В этих условиях можно судить о состоянии сети и определять тип передаваемых данных лишь по косвенным признакам.

Хороших результатов при анализе сетевой активности можно добиться, используя метод энтропийного анализа передаваемых по сети информационных пакетов [1, 2]. В данном случае во внимание берётся тот факт, что данные по сети передаются небольшими порциями (пакетами), каждый из которых снабжён набором служебных меток (полей заголовков TCP/IP). Совокупность этих меток для группы пакетов в потоке имеют свою уникальную статистическую характеристику, которую удобно измерять при помощи информационной энтропии, предложенной Кл. Шенноном.

Первые работы в направлении исследования энтропийных характеристик сетевых потоков были связаны с выявлением эпидемий сетевых червей [1]. Экспериментальный анализ показал существенное отличие энтропийных характеристик потоков данных, передаваемых по сети в период распространения сетевых червей и в период их отсутствия [2]. На основе этого были сделаны предположения о возможности использования энтропийного анализа для детектирования ранних стадий распространения сетевых червей. На практике предложенные реализации определяли величину энтропии по степени сжатия заголовков информационных пакетов известными алгоритмами архивации, такими, как zip, lzip и им подобными [2]. Этот способ весьма трудоёмок по сравнению с непосредственным подсчётом и даёт опосредованное представление об истинной величине

не энтропии. Однако компрессия данных имеет одно достоинство: получаемая в результате величина степени сжатия оцениваемых данных лежит в диапазоне от 0 до 1 независимо от прочих факторов оценки. При прямом расчёте энтропии получается положительная величина, максимальное значение которой зависит от количества реализаций случайных событий, которые могут встретиться в потоке. Авторами было показано, что можно соотнести полученное значение энтропии со степенью сжатия потока и привести его к диапазону от нуля до 1, поделив его на теоретически максимально возможное значение энтропии для конкретной выборки, не теряя при этом возможности идентификации программного обеспечения, генерирующего исследуемые сетевые потоки. Данная модификация в сравнении с предыдущими предложенными авторами реализациями [3] не повлияла на скорость обработки данных, однако позволила существенно расширить список классов детектируемого программного обеспечения, отличного от сетевых червей. Практическое измерение авторами скорости обработки потоков данных методами, предложенными в работе А. Wagner [2], и непосредственного подсчёта энтропии, используемого авторами, показало превосходство последнего в 5 раз и более. Чтобы исключить влияние на эксперимент сторонних факторов, обрабатывались заранее подготовленные потоки данных размером порядка нескольких сотен мегабайт, размещённые в оперативной памяти компьютера, на котором проводились эксперименты.

Предложенный подход можно обобщить для установления принадлежности сетевых потоков не только сетевым червям, но и другим программам, таким, как клиенты пиринговых сетей, браузеры, сетевые сканеры и так далее [4]. Каждый класс программного обеспечения в таком списке имеет свои, присущие только ему энтропийные характеристики генерируемых потоков, по которым его можно легко выделить среди остальных. Чтобы воспользоваться данным методом установления принадлежности сетевых потоков, необходимо заранее подготовить базу энтропийных характеристик потоков, генерируемых идентифицируемым программным обеспечением.

На основе предложенного авторами метода выполнена программная реализация дифференциации сетевых потоков, используемая для мониторинга состояния сети радиотехнического факультета Томского государственного университета [3]. Программа успешно помогает выявлять аномалии состояния сети и ликвидировать большой класс проблем на ранней стадии их возникновения. Сам метод энтропийного анализа характеристик сетевых потоков пригоден для решения широкого круга прикладных и фундаментальных задач, таких, как соблюдение корпоративных политик использования сети, обнаружение нежелательных сетевых программ, обеспечение приоритета передачи данных определённого программного обеспечения, выявление скрытой сетевой активности, выявление попыток сканирования сети или отдельных компьютеров, выявление факта распространения в сети компьютерных червей, обнаружение удалённого управления компьютерами без ведома пользователей, выявление источников массовых нежелательных рассылок, анализ состояния сети и повышение её управляемости.

**Сравнение скорости обработки данных алгоритмов
Izop и Entromer**

Алгоритм	Скорость, МБ/с	Энтропия $H/H_{\max}$
Izop	38	1
Непосредственный расчёт	270	0.97

СПИСОК ЛИТЕРАТУРЫ

1. Kim J., Radhakrishnan S., and Dhall S. K. // Computer Communications and Networks. ICCCN 2004: Proc. 13th International Conference on. – 2004. – P. 495–500.
2. Wagner A. Entropy-Based Worm Detection for Fast IP Networks.: Theses ... Ph.D. on techn. sc. / Swiss Federal Institute of Technology Zurich. – Zurich, 2008. – 115 p.
3. Булахов Н.Г., Спиридонов. А.А., Минаев А.И., Крестьянников А.С. // Изв. вузов. Физика. – 2012. – Т. 55. – № 9/2. – С. 325–326.
4. 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;
Медведев Михаил Дмитриевич, студент;
Кузнецов Владимир Сергеевич, студент.

*N.G. BULAKHOV, M.D. MEDVEDEV, V.S. KUZNETSOV***PRACTICAL APPLICATION OF ENTROPY ANALYSIS METHOD OF NETWORK FLOWS**

Consider the options and features of the method of the entropy analysis of parameters of network packet flows for applications computer network management and process descriptions in them.

Keywords: *entropy analysis, computer networks, network flows.*