УДК 004.051

Шыхалиев Р.Г.
Институт Информационных Технологий НАНА, Баку, Азербайджан
ramiz@science.az

ОБ ОДНОМ ПОДХОДЕ К МОДЕЛИРОВАНИЮ МОНИТОРИНГА КОМПЬЮТЕРНЫХ СЕТЕЙ

Мониторинг играет важную роль для эффективного управления компьютерными сетями (КС). При этом основными элементами мониторинга КС являются потоки сетевых трафиков. В статье на основе существующих моделей потоков данных предлагается модель мониторинга КС.

Ключевые слова: *мониторинг компьютерных сетей, IP-трафик, потоки IP-пакетов, IP-адреса, модели потоков данных.*

Введение

Мониторинг и анализ сетевых трафиков компьютерных сетей (КС) имеют жизненно важное значение для их эффективного управления. Однако увеличение скорости связи и объемов сетевых трафиков делает мониторинг и анализ сетевых трафиков КС очень сложными. Одними из основных проблем являются онлайн-мониторинг и интерактивный анализ сетевых трафиков КС.

Сетевые трафики современных КС состоят из больших потоков IP-пакетов, которые имеют очень большие объемы (терабайты или петабайты). При этом сбор и обработка всех сетевых трафиков очень важны, так как заранее неизвестно, какие части сетевых трафиков будут содержать информацию об интересующих администраторов событиях. Поэтому для того, чтобы иметь полную картину функционирования КС, необходимо осуществлять сбор всех сетевых трафиков, иначе могут быть пропущены важные события сети. Хотя сегодня сбор и хранение большого объема сетевых трафиков не являются проблемой, все же извлечение ценной информации из собранных сетевых данных вызывает трудности. Это связано с тем, что способность средств сбора и хранения данных растет более быстрыми темпами, чем умение их анализировать.

При мониторинге КС анализ сетевых трафиков в основном происходит на двух уровнях. На первом анализ сетевого трафика осуществляется на пакетном уровне [1]. А на втором уровне анализируются сами потоки трафиков [2, 3]. Однако большинство используемых сегодня систем сетевого мониторинга и анализа не позволяет обрабатывать потоки сетевых трафиков и проводить мониторинг КС в режиме реального времени. В таких системах результаты мониторинга в основном анализируются «ручным» способом или на основе лог-анализов. Вместе с тем непрерывное и быстрое поступление потоков данных из сетевого трафика в систему мониторинга приводит к накапливанию большого объема данных, который должны обработать сетевые администраторы. В результате увеличивается время, которое должны потратить администраторы сетей для анализа всех сетевых данных, что приводит к снижению эффективности принятых решений по управлению КС. В таких условиях с помощью традиционных стандартных анализаторов сети трудно добиться понимания природы сетевых трафиков. Поэтому для мониторинга КС предлагается использовать модели обработки потоков данных.

Целью данной статьи является создание модели мониторинга КС. Для этого предлагается использовать существующие модели потоков данных. Это позволит осуществлять мониторинг КС в режиме реального времени и решать проблемы, связанные с хранением большого объема данных.

Анализ трафиков КС и задачи мониторинга

Исследования сетевых трафиков показали, что они представляют собой сложный динамический процесс и являются суперпозицией потоков с множественными взаимосвязанными характеристиками, которые генерируются различными протоколами. Во-первых, это трафики, связанные с управлением КС (например, трафик инициализации клиентов, серверный трафик и т.д.), которые генерируются периодически. Во-вторых, это трафики сетевых сервисов, приложений (например, DNS, FTP, запросы WINS, ARP, сеанс NetBIOS, HTTP, P2P, SMTP, POP3, Telnet и т.д.) и протоколов, которые составляют основную часть сетевого трафика КС [4].

Как известно, IP-протокол является универсальным протоколом для любого типа приложений, используемых в КС, и вся нагрузка по транспорту трафиков ложится на него. Также известно, что основными транспортными протоколами, которые работают на IP, являются TCP и UDP. Сетевой трафик КС в основном состоит из TCP-трафиков, что является основной особенностью IP-трафика. При этом, если рассмотреть более детально, структура сетевого трафика КС состоит из IP-пакетов. Все трафики в КС, например веб-трафики, трафики приложений, состоят из серии IP-пакетов. Каждый IP-пакет состоит из заголовка и отправляемой информации [5]. В свою очередь заголовок пакета состоит из полей с информацией об IP-адресе источника (отправителя), IP-адресе назначения (получателя), времени жизни пакета, типе сервиса, типе протокола, номере порта и т.д. При этом наиболее важными полями заголовка IP-пакета являются IP-адреса источника и назначения. Они используются для того, чтобы идентифицировать хосты отправителя и получателя. В КС каждый хост при соединении с другими хостами представляется уникальным идентификатором, состоящим из IP-адреса и номера порта. А на более высоком уровне агрегации имеются потоки IP-трафиков (веб-трафик, трафик приложений и т.д.), представляющие из себя наборы IP-пакетов с некоторыми одинаковыми ключевыми атрибутами, например, IP-адреса источника и назначения, номер порта, тип протокола и т.д.

В зависимости от целей мониторинга КС могут быть рассмотрены различные задачи, например, такие как:

- определение количества IP-адресов (различных), использующих в течение дня заданный канал связи в текущем потоке;
- определение количества отправленного веб-трафика из определенного диапазона IP-адресов;
- определение самых интенсивных потоков IP-трафиков;
- определение количества потоков, содержащих только один пакет (т.е. необычные потоки);
- определение количества пар коррелированных связей между потоками IP-трафиков;
- определение для каждого IP-адреса источника в определенном временном интервале объема веб-трафика и количества пакетов, то есть определение интенсивности веб-трафика для каждого IP-адреса источника и т.д.

Модели потока данных

Поток данных – это последовательность упорядоченных (имеющих временные отметки или не имеющих) элементов, непрерывно поступающих в реальном масштабе времени. Обычно потоки данных поступают очень быстрыми темпами и невозможно упорядочить их поступление, а из-за ограниченности памяти становится трудным хранение потоков в целом.

Так как элементы могут поступать в отдельности, поток данных может быть смоделирован как последовательность списков элементов [6]. Отдельные элементы потока могут принимать форму реляционных кортежей или спецификации объектов.

В модели потока данных к входным данным, которые необходимо обработать, нет возможности произвольного доступа из диска или памяти, и они поступают как последовательность непрерывных потоков данных. Обычно при произвольном способе доступа к дискам или памяти для чтения/записи произвольного блока данных не требуется последовательного просмотра блоков, начиная с самого первого, при этом для доступа к разным блокам данных тратится почти одинаковое время.

Модель потоков данных имеет некоторые отличия от обычной модели данных с реляционными отношениями:

− элементы данных в потоке поступают в реальном масштабе времени;

− система не имеет никакого контроля над упорядочением элементов потоков данных;

− потоки данных не имеют ограничения в размере;

− после обработки элемента потока данных он исключается из потока или архивируется, при этом для дальнейшего восстановления его нужно сохранить в памяти, которая имеет намного меньший размер, чем размер потока данных в целом.

Вместе с тем при обработке информации в модели потока данных не исключается наличие данных, имеющих обычные реляционные отношения.

У потоков данных имеется несколько моделей, и формально потоки данных можно описать следующим образом. Входной поток $s_1, s_2, ...$ поступает последовательно, элемент за элементом, и описывает основной сигнал S, который является одномерной функцией – $S:[1...N]\to R$. При этом вход может состоять из нескольких потоков или многомерных сигналов. В зависимости от того, как элементы s_i описывают S, модели потока данных отличаются:

− модель временных рядов. Каждый s_i-ый элемент равняется $S[i]$, и они поступают в порядке возрастания значения i. Эта модель подходит для временных рядов данных, например, при наблюдении за объемом сетевого трафика КС каждые 10 минут и т.д. В каждом таком периоде времени мы наблюдаем последующие новые данные;

− модель кассового аппарата. Здесь s_i-ые элементы возрастают до $S[j]$-х. Допустим, что $s_i = (j, I_i), I_i \geq 0$, это означает, что $S_i[j] = S_{i-1}[j] + I_i$, где S_i − это состояние сигнала после наблюдения i-го элемента в потоке. Как в кассовом аппарате, множество s_i-е могут со временем возрастать до определенного $S[j]$. Эта модель является самой популярной моделью потока данных. Она подходит для мониторинга в КС IP-адресов, которые обращаются к веб-серверу и отправляют пакеты в сеть и т.д. Так как в КС одни и те же IP-адреса могут несколько раз получить доступ к веб-серверу или отправлять множество пакетов в сеть в течение долгого времени;

− модель турникета. Здесь s_i-ые элементы обновляются на $S[j]$-х. Допустим, что $s_i = (j, I_i)$, это означает, что $S_i[j] = S_{i-1}[j] + I_i$, где S_i − это сигнал после наблюдения i-го элемента в потоке и I_i может быть положительным или отрицательным. Эта модель является самой общей моделью потока данных, которая позволяет исследовать динамические ситуации, где элементы непрерывно поступают в систему и покидают ее.

Эти модели более подробно описаны в работе [7].

Модели мониторинга КС

Для решения задач мониторинга КС, указанных в разделе 2, используем модели потоков данных. На основе модели потока данных формально опишем задачу определения количества IP-адресов (различных), использующих в течение дня заданный канал связи в текущем потоке. Для решения этой задачи более подходящей является модель кассового аппарата.

Пусть поток $p_1, p_2, ...$ представляет собой последовательность IP-пакетов в заданном канале связи с пакетами p_i, имеющими IP-адреса источников s_i. Пусть $P[0....N-1]$ является количеством пакетов, переданных IP-адресом источника i, и для $0 \le i \le N-1$ в начале дня инициализируются нулями. А при каждом поступлении пакета p_i в $P[s_i]$ добавляется единица. Число различных IP-адресов, которые использовали данный канал связи в течение дня, может быть определено путем подсчета количества ненулевых $P[i]$-х в любое время.

Также очень важным является определение количества различных IP-адресов, использующих заданный канал связи в текущие моменты времени. Для этого используем модель турникета. Формально это определяется так: в любое время t проводится подсчет IP-адресов s_i, которые входят в некоторый поток f_i. При этом для идентификации в потоке IP-пакетов первого и последнего пакетов используется информация идентификации порядка пакетов, то есть номера пакетов. Это позволяет определить начало и конец потока IP-пакетов. Пусть $P[0....N-1]$ является количеством потоков, которые в текущий момент времени входят в IP-адреса источников i и для $0 \le i \le N-1$ в начале дня инициализируются нулями. Если пакет p_i находится в начале потока и s_j является источником пакета p_i, то в $P[s_j]$ добавляется единица, а если пакет p_i находится в конце потока и s_j является источником пакета p_i, то из $P[s_j]$ вычитывается единица, в противном случае никаких операций не проводится. Подсчет количества различных IP-адресов, использующих данный канал связи в текущие моменты времени, может быть осуществлен путем определения количества ненулевых $P[i]$-х в любое время.

Аналогично другие вышеуказанные задачи мониторинга КС могут быть формализованы в терминах моделей потока данных и вычислены соответствующие функции.

Заключение

Как известно, трафики современных КС состоят из больших потоков IP-пакетов и имеют очень большие объемы (терабайты или петабайты). В результате этого мониторинг и анализ сетевых трафиков КС в режиме реального времени становятся очень сложными.

Традиционно используемые системы сетевого мониторинга и анализа не позволяют обрабатывать потоки сетевых трафиков в режиме реального времени и проводить эффективный мониторинг КС. Поэтому для мониторинга КС предлагается использовать модели потоков данных и обосновывается выбор различных моделей потока данных в зависимости от целей мониторинга КС.

Литература

1. Dainotti A., Pescape A/ and Ventre G. A Packet-level Characterization of Network Traffic / 11th International Workshop on Computer-Aided Modeling, Analysis and Design of Communication Links and Networks, Trento, 2006, June 8–9, pp. 38–45.
2. Lakhina A., Papagiannaki K. et al. Structural Analysis of Network Traffic Flows / SIGMETRICS/Performance'04, 2004, June 12–16, New York, USA, 2004, pp. 61–72.
3. Sierszen A., Sturgulewski L. Traffic Analyzer Based on Data Flow Patterns // Automatyka, Tom 15, Zeszyt 3, 2011, pp. 693–702.
4. Шыхалиев Р.Г. Анализ и классификация сетевого трафика компьютерных сетей // Проблемы информационных технологий, 2010, № 2, с. 15–23.
5. http://www.ietf.org/rfc/rfc791.txt
6. Gilbert A.C., Kotidis Y., Muthukrishnan S. and Strauss M.J. One-Pass Wavelet Decompositions of Data Streams / IEEE Transactions on Knowledge and data Engineering, May/June 2003, Vol. 15, No. 3, pp. 541–554.
7. Tucker P., Maier D., Sheard T., Fegaras L.. Enhancing relational operators for querying over punctuated data streams. 2002. www.cse.ogi.edu/dot/niagara/pstream/punctuating.pdf.

UOT 004.048

Şıxəliyev Ramiz H.
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
ramiz@science.az

Kompüter şəbəkələrinin monitorinqinin modelləşdirilməsinə bir yanaşma haqqında

Monitorinq kompüter şəbəkələrinin (KŞ) effektiv idarə edilməsi üçün vacib rol oynayır. Bununla yanaşı KŞ-in monitorinqinin əsas elementləri şəbəkə trafiki axınlarıdır. Məqalədə verilənlər axını modelləri əsasında KŞ-in monitorinqi modeli təklif olunmuşdur.

Açar sözləri: *kompüter şəbəkələrinin monitorinqi, IP-trafik, IP-paketlər axını, IP-ünvanlar, verilənlər axını modelləri.*

Ramiz H.Shikhaliyev
Institute of Information Technology of ANAS, Baku, Azerbaijan
ramiz@science.az

On one approach to modeling of the computer networks monitoring

Monitoring plays an important role in efficient management of computer networks (CN). The network traffic streams are basic elements of the CN monitoring. The paper proposes a model for monitoring of CNs on the basis of existing data streams models.

Key words: *computer networks monitoring, IP-traffic, IP-packets stream, IP-addresses, data stream models.*