О Методах Сбора и Хранения Сетевого Трафика Компьютерных Сетей

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

Аннотация **− процесс сбора и хранения сетевого трафика компьютерных сетей (КС) является одним из основных этапов процесса мониторинга. Однако в современных КС процесс сбора и хранения полного сетевого трафика представляет собой очень сложную проблему. Так как с ростом скорости и масштаба КС растут и объемы сетевого трафика, который в день может потребовать петабайты объемов памяти для хранения. Существуют различные методы сбора и хранения сетевых данных, которые при правильном выборе могут существенно снизить объем собранных данных, следовательно, и объем требуемых для хранения данных. Исходя из этого, в статье рассматриваются подходы к решению вопросов сбора и хранения данных, а также проблемы, имеющиеся в этой области.**

Ключевые слова − компьютерные сети, мониторинг, сетевой трафик, сбор сетевого трафика, хранение сетевого трафика.

I. ВВЕДЕНИЕ

Сегодня компьютерные сети, особенно Интернет, стали глобальной инфраструктурой, обеспечивающей повсеместно доступные, интерактивные и безопасные сервисы. Для того чтобы обеспечить высокий уровень QoS (Quality of Service) этих сервисов, необходима эффективная инфраструктура мониторинга. При этом самым подходящим методом для инфраструктуры мониторинга КС является пассивный мониторинг [1], который позволит определить общее состояние КС и безопасность, а также уровень QoS предоставляемых сервисов и т.д. Для этого необходимо постоянно осуществлять сбор, хранение и анализ большого объема данных, то есть постоянно осуществлять сбор, хранение и анализ сетевого трафика. Однако это представляет из себя очень сложную задачу, в особенности при мониторинге больших КС. Так как с повышением скорости и масштабов КС растет и объем сетевого трафика, который в день может измеряться петабайтами. Вместе с тем для того, чтобы провести всеобъемлющий анализ состояния и безопасности КС, необходимо произвести сбор и хранение всей информации о трафике. Например, при мониторинге безопасности КС сбор всех сетевых пакетов необходим для обнаружения вредоносных активностей или определения вирусных атак (например сетевыми червями) и т.д.

Известно, что при пассивном мониторинге КС собираются большие объемы данных мониторинга, что приводит к возникновению проблем, связанных с их хранением и уменьшением эффективности анализа. Поэтому, наряду с необходимостью разработки новых методов анализа больших объемов сетевых трафиков, актуальной проблемой является разработка новых подходов для сбора и хранения больших сетевых трафиков для мониторинга КС. При этом очень важным является сокращение размерности признакового пространства сетевого трафика, который используется для мониторинга КС [2]. Исходя из этого, в статье рассматриваются подходы к решению вопросов сбора и хранения, а также проблемы, имеющиеся в этой области.

II. МЕТОДЫ СБОРА СЕТЕВОГО ТРАФИКА

Существующие сегодня средства мониторинга осуществляют сбор различных типов и объемов информации. При этом имеются три основных метода сбора трафика, которые имеют различные требования к объему памяти: сбор всех пакетов; сбор сетевого потока и так называемый сбор расширенного потока.

Целью сбора всех пакетов является сбор всего сетевого трафика, который генерируется компьютерами и устройствами КС, при котором осуществляются сбор и хранение данных заголовка каждого пакета и передаваемой в пакетах информации. Другими словами, производятся сбор, обработка и хранение копии каждого пакета трафика для последующего анализа. Эти собранные данные обеспечивают аналитиков полной информацией о трафике: полной информацией заголовков пакетов и передаваемой в пакетах информацией. Следовательно, такой метод сбора данных мониторинга может быть наиболее универсальным, так как большой объем информации может интенсивно храниться и обрабатываться [3].

Сетевой поток определяется как множество IP-пакетов, проходящих через точку наблюдения в сети в течение определенного интервала времени. Все пакеты, принадлежащие к определенному потоку, имеют набор общих свойств. Требования к потокам IP-пакетов определены в RFC 3917 [4]. Согласно приведенному определению, сетевой поток представляет из себя потоки сетевых пакетов, для которых выполняются следующие условия:

- происходят в течение одного и того же периода времени;
- имеют один и тот же адрес источника и номер порта;

- имеют один и тот же адрес назначения и номер порта;
- используют один и тот же протокол.

При этом, если не рассматривать передаваемую в пакетах информацию и информацию некоторых полей заголовков пакетов, а также объединить некоторые пакеты, то уменьшится объем данных, что приводит к уменьшению требуемой памяти для хранения потоков. Однако это вызывает понижение качества анализа сетевого трафика [5].

Сбор расширенного потока включает в себя сбор всех пакетов и сетевого потока. При этом к информации потока добавляется информация, взятая непосредственно из заголовков пакетов или из передаваемой в пакетах информации. Вместе с тем расширенный поток также может содержать дополнительную информацию о каком-то внешнем источнике, например, о географическом расположении IP-адресов источника и назначения. Поэтому некоторые решения сбора расширенного потока рассматривают эту информацию как метаданные [6].

Большая часть современных исследований в области сбора сетевого трафика посвящена вопросам сбора пакетов в скоростных сетях с минимальной потерей данных и сжатием данных после сбора, то есть снижению объемов. Например, в работах [7, 8] авторы соответственно обсуждают вопросы преобразования данных для их эффективного хранения и обработки и мониторинга в облаке (Cloud). В работах [9, 10] авторы предлагают подход к разработке приложений по сбору данных в скоростных сетях, основанный на стандартных аппаратных средствах. А в работе [11] для полного анализа сетевого трафика авторы предлагают метод агрегации потоков.

III. МЕТОДЫ ХРАНЕНИЯ СЕТЕВОГО ТРАФИКА

Другой проблемой эффективного анализа сетевого трафика является хранение собранных данных, которые должны сохраняться достаточно долго и надежно, чтобы при необходимости аналитики могли иметь доступ к ним. При этом, в зависимости от места и способа хранения данных, может существенно изменяться требуемый объем памяти для хранения, а также возникать проблемы, связанные с администрированием и обслуживанием и т.д. Вместе с тем данные могут быть сохранены локально в организации, облаке или другом внешнем хранилище и могут быть использованы различные способы хранения данных, такие, как файлы (например, лог-файлы); базы данных и их комбинации. Каждый из этих способов имеет собственные аспекты.

Обычно в большинстве организаций КС производят сбор сетевого трафика в нескольких точках. Поэтому очень важен выбор места физического расположения собранных данных. Например, централизованное хранение всех собранных данных в одном месте может упростить управление и анализ данных, однако требует передачи данных в центр, что приводит к неэффективному использованию полосы пропускания каналов передачи сети. Также такой способ хранения данных нецелесообразен с точки зрения безопасности, так как при компрометации хранилища может произойти несанкционированное изъятие данных. Альтернативой централизованного хранения данных является распределенное хранение данных, однако при таком подходе хранения усложняется процесс анализа данных, а также администрирования и обслуживания. Одним из видов распределенного хранения данных можно считать облачное хранение данных [12, 13], который может также осуществлять и сбор данных.

Большинство средств сбора сетевого трафика записывают полученные данные в файлы (лог-файлы) и обычно имеют собственные форматы файлов. При этом очень важно знать формат хранимого файла, чтобы без затруднения организовать передачу данных между приложениями сбора и анализа данных, так как большинство из них поддерживают определенные форматы файлов. Однако имеются некоторые общие форматы (например, pcap), которые поддерживаются большинством приложений сбора и анализа данных. Вместе с тем форматы создаваемых файлов могут определить необходимые объемы для хранения файлов. Несмотря на то, что в малых объемах данных различия между форматами не существенны, в больших объемах данных выбор того или иного формата является существенным. Уменьшение объема памяти, необходимого для хранения данных, также может быть достигнуто сжатием данных, которое может сделать хранение и анализ данных более эффективными. Эффективный алгоритм сжатия может не только уменьшить дисковое пространство, необходимое для хранения данных, но также уменьшить время, требуемое для извлечения данных из этого диска. Например, алгоритм сжатия данных lzo1x может уменьшить размер записей примерно на 50% [14].

Некоторые средства сбора данных для хранения данных могут использовать базы данных. Приложения, которые поддерживают только файлы, могут хранить их в базе сами или с помощью приложения анализа данных. Аналитик это может сделать вручную. Вместе с тем при использовании базы данных для хранения сетевого трафика необходимо учитывать ожидаемый размер и количество записей и свойства базы данных, ограничивающих общий размер базы данных, размер записи и т.д. Учитывая то, что реляционные базы данных не так масштабируемы, как при хранении данных в виде файлов. Для решения проблемы масштабируемости могут быть использованы NoSQL базы данных, такие, как Hadoop [15].

ЗАКЛЮЧЕНИЕ

Сегодня мониторинг сетевого трафика КС является одним из основных средств обеспечения их нормальной работы и безопасности. При этом сбор и хранение являются основой мониторинга. Для получения полной информации о деятельности КС необходимы постоянный и полный сбор и хранение сетевого трафика, что позволяет своевременно и эффективно реагировать на отказы в работе и инциденты безопасности. Однако это требует постоянного сбора и хранения большого объема данных

мониторинга, что может требовать огромных объемов памяти, а также снизить эффективность анализа собранных данных. Другой проблемой сбора и хранения большого объема сетевого трафика, на наш взгляд, является неправильный выбор соответствующих методов сбора и хранения данных в зависимости от задач мониторинга.

В статье были проанализированы существующие методы сбора и хранения сетевого трафика, а также проблемы, имеющиеся в этой области.

В результате анализа можно сделать вывод, что при сборе сетевого потока или расширенного потока требуется гораздо меньше памяти для хранения. Проведенный в статье анализ методов сбора и хранения сетевого трафика поможет администраторам КС в соответствии с задачей мониторинга выбрать необходимый метод.

ЛИТЕРАТУРА

[1] Şıxəliyev R.H. Kompüer şəbəkələrinin monitorinqi üsulları və vasitələri haqqında // İnformasiya cəmiyyəti problemləri, №2(4), səh. 61-70, 2011.

[2] Шыхалиев Р.Г. Об одном методе сокращения размерности анализируемых признаков сетевых трафиков, используемых для мониторинга компьютерных сетей // Телекоммуникации. - N: 06. - с. 44-48, 2011.

[3] Bejtlich Richard., Why Collect Full Content Data? http://taosecurity.blogspot.com', 2012

[4] Quittek J., Zseby T., Claise B., Zander S., RFC 3917: Requirements for IP Flow Information Export (IPFIX). Internet Engineering Task Force, 2004. http://tools.ietf.org/html/rfc3917

[5] RFC 7011, Specification of the IP Flow Information Export (IPFIX) Protocol, a standardized network flow format, provides a more technical definition of flow. http://tools.ietf.org/search/rfc7011

[6] National Information Standards Organization (NISO). Understanding Metadata. NISO, 2004.

[7] Aceto G., Botta A., Pescape A., Westphal C. Efficient Storage and Processing of High-Volume Network Monitoring Data, IEEE Transactions on Network and Service Management, vol. 10, issue 2, pp. 162-175, 2013.

[8] Aceto G., Botta A., de Donato W., Pescape A., Cloud Monitoring: A Survey, Computer Networks vol.57, issue 9, pp. 2093-2115, 2013.

[9] Deri L., Cardigliano A., Fusco F., 10 Gbit Line Rate Packet-to-Disk Using n2disk, Proceedings IEEE INFOCOM. Turin, Italy, Apr. IEEE, pp. 3399-3404, 2013.

[10] Banks D., Custom Full Packet Capture System, SANS, 2013.

[11] Francois J., State R., Engel T., Aggregated Representations and Metrics for Scalable Flow Analysis, IEEE Conference on Communications and Network Security (CNS). Washington, D.C., pp. 478-482, 2013.

[12] Sivashakthi T., Prabakaran N., A Survey on Storage Techniques in Cloud Computing, International Journal of Emerging Technology and Advanced Engineering, Vol. 3, Issue 12, pp. 125-128, 2013.

[13] Spoorthy V., Mamatha M., Santhosh Kumar B., A Survey on Data Storage and Security in Cloud Computing, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.6, 2014, pp. 306-313.

[14] Software Engineering Institute, Carnegie Mellon University. SiLK FAQ https://tools.netsa.cert.org/silk/faq.html (2014).

[15] Hadoop, http://hadoop.apache.org/