ТЕХНОЛОГИЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА И АВТОМАТИЧЕСКОЙ ОПТИМИЗАЦИИ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В.А. Спиридонов, *Московский технический университет связи и информатики, spiridonov-valeriy@bk.ru.*

УДК 004.8:004.415.53

Аннотация. Деятельность центров обработки данных (ЦОД) по эксплуатации и техническому обслуживанию сети сосредоточена в основном на проверке рабочих состояний устройств. Инженеры по эксплуатации и техническому обслуживанию определяют, как работают службы и пропускную способность ЦОД, проверяя рабочее состояние устройств. Однако этот метод не может отражать реальное состояние передачи бизнес-данных, поэтому инженеры не могут всесторонне оценить общие условия работы предприятий. В данной статье предлагается использование технологии анализатора портов инкапсулированного удаленного коммутатора (*Encapsulated remote switch port analyzer, ERSPAN*) для анализа нагрузки сети. Это позволяет инженерам по эксплуатации и техническому обслуживанию всесторонне оценивать состояние обслуживания в ЦОД и формировать тесно связанную модель корреляции между сетями и службами посредством сквозного визуализированного моделирования, обеспечивая техническую поддержку для оптимизации ЦОД и раннего предупреждения о сетевых рисках.

Ключевые слова: анализатор; анализ; сеть; трафик, протокол.

TECHNOLOGY OF NETWORK TRAFFIC ANALYSIS AND AUTOMATIC OPTIMIZATION BASED ON ARTIFICIAL INTELLIGENCE

Vladimir Spiridonov, *Moscow technical university of communications and informatics.*

Annotation. The network operation and maintenance activities of data processing centers are mainly focused on checking the operating states of devices. Operation and maintenance engineers determine how the services and throughput of data centers work by checking the operational status of the devices. However, this method cannot reflect the actual state of business data transmission, so engineers cannot comprehensively assess the general working conditions of enterprises. This article proposes an approach to using the technology of the Encapsulated Remote Switch Port Analyzer (*ERSPAN*) for network load analysis. This allows operation and maintenance engineers to comprehensively assess the state of service in data centers and form a closely related correlation model between networks and services through end-to-end visualized modeling, providing technical support for data center optimization and early warning of network risks.

Keywords: analyzer; analysis; network; traffic; protocol.

Введение

Сетевой анализатор создается для сбора статистики о размере потока данных в сеансе протокола управления передачей (*Transmission Control Protocol, TCP*) на основе порядкового номера в заголовках *TCP* пакетов синхронизации номеров последовательности (*Synchronize sequence numbers, SYN*) и флага указания на завершение соединения (*final, FIN*).

Трафик отправителя представляет собой порядковый номер в пакете *FIN-ACK*, а также порядковый номер в пакете *SYN*, который отправляется от отправителя.

Трафик от получателя представляет собой порядковый номер в пакете *FIN-ACK* (от англ. *ACK – Acknowledgement field is significant,* номер подтверждения), который отправляется от получателя и порядковый номер в пакете *SYN-ACK,* который отправляется от получателя.

Порядковый номер в заголовке *TCP* представляет собой 32-разрядное поле. По мере установления сеанса *TCP* значение порядкового номера может учитываться в обратном порядке. Если обратный подсчет происходит только один раз во время установления сеанса *TCP,* алгоритм сбора статистики может определить это во время вычисления и автоматически скорректировать значение. Если значение порядкового номера подсчитывается в обратном порядке несколько раз во время сеанса *TCP,* алгоритм не может идентифицировать отправителя и получателя, что приводит к ошибке в статистике трафика.

Восстановление пути пересылки *TCP*-пакета происходит следующим образом. Каждый раз, когда *TCP*-пакет поступает на сетевое устройство определенного уровня, значение поля времени жизни (*Time to live, TTL*) пакета данных в протоколе в заголовке *IP*-пакета уменьшается на единицу. Основываясь на этом принципе, алгоритм сетевого анализа сначала сопоставляет все собранные *TCP*-пакеты на основе содержимого внутреннего пакета и определяет, что *TCP*-пакеты, собранные с маршрутизаторов, принадлежат одному и тому же *TCP*-сеансу. Затем анализатор сортирует пакеты в порядке убывания на основе внутренних и внешних значений *TTL* с последующим выполнением вычислений на основе определенных правил сопоставления и идентификации для восстановления пути пересылки исходного *TCP*-пакета.

Вычисление задержки *TCP*-пакетов

Пакеты в потоке данных с зеркальным отображением *ERSPAN* не могут содержать информацию о временной метке, поэтому время, когда поток данных TCP достигает каждого устройства, фактически записывается временной меткой, вставленной анализатором после того, как пакеты в потоке данных зеркально отображаются на сборщике. Когда анализатор выполняет вычисление восстановления пути, он вычисляет задержку *TCP*-пакета для каждого перехода на основе временной метки, поставленной им же.

Для сбора пакетов *TCP SYN* для маршрутизаторов настраивается зеркальный поток пакетов данных *ERSPAN*. В соответствии с меткой времени, вставленной анализатором, когда он получает пакеты *TCP SYN,* зеркально отраженные от маршрутизаторов, задержка для каждого перехода составляет разность задержек передачи данных между соответствующими маршрутизаторами. Однако в реальной сетевой среде, поскольку пути передачи и анализаторы для обработки потоков данных различаются, а в локальном кэшировании маршрутизаторов вдоль пути существуют задержки, время, когда три зеркальных пакета поступают на анализатор, и порядок, в котором исходные пакеты достигают каждого маршрутизатора, различны. В результате временная метка, вставленная анализатором, может не быть корректной, что дополнительно приводит к разнице в ошибках между задержкой, полученной путем вычисления, и фактической задержкой.

Чтобы предотвратить эту проблему, поток данных с зеркальным отображением *ERSPAN* должен быть настроен для маршрутизаторов в направлении от приемника для сбора пакетов *SYN-ACK*. Анализатор сетевого ввода-вывода вычисляет разницу между временем, когда устройство получает *SYN-ACK*, и временем, когда оно получает *SYN*. Если значение времени передачи маршрутизатора (*Router transmission time, RTT*) больше или равно 0, то большее

значение, полученное ранее, является ненормальным и должно рассматриваться как недопустимое значение. Если значение *RTT* меньше 0, то можно использовать большее значение, полученное ранее.

Идентификация приложения TCP. На основе информации из пяти кортежей во внутренних пакетах *TCP*, наряду с номером хоста и портом прикладного уровня, можно точно идентифицировать конкретное приложение. Алгоритм идентификации приложения анализатора требует, чтобы пользователь сначала ввел *IP*-адрес и имя приложения, которое необходимо идентифицировать. При обработке потоков данных *TCP* анализатор автоматически сопоставляет вводимую пользователем информацию с информацией из пяти кортежей внутренних *TCP*-пакетов, а также номером хоста и портом прикладного уровня, чтобы точно определить, какое приложение отправило поток данных *TCP*. Анализатор также использует алгоритмы искусственного интеллекта (ИИ) для анализа данных собранного сетевого трафика, чтобы активно определять соответствующие отношения между пятью кортежами пакетов, информацией прикладного уровня и приложением, подлежащим идентификации, формируя таким образом автономную базу данных функций. При идентификации приложения, отправившего *TCP*-пакеты, автоматически проверяется база данных, что эффективно позволяет избежать ошибок идентификации, вызванных неправильным вводом пользователем.

Внутриполосная сетевая телеметрия (*In-band network telemetry, INT)* – технология мониторинга сети, которая собирает данные с устройств. Оборудование отправляет собранные данные в анализатор, обеспечивает функцию сбора данных в режиме реального времени и высокой скорости и достигает цели мониторинга производительности сетевого оборудования и работы сети.

Благодаря технологии *INT* можно отслеживать входные и выходные порты и информацию об очереди каждого устройства на пути пересылки сообщений, информацию о временных метках устройств ввода и вывода, информацию о перегрузке очереди и т.д. При последнем переходе обнаружения пути отслеживаемые данные инкапсулируются заголовком протокола пользовательских датаграмм (*User datagram protocol, UDP*) и *IP*-заголовком и перенаправляется анализатору трафика. Наконец, с помощью программного обеспечения (ПО) для управления сетью, развернутого на анализаторе, проверяются данные мониторинга и извлекается полезная информация.

С помощью технологии *INT* можно представить точный путь пересылки по сети и задержку потока приложений, а также напрямую определить работоспособность приложений в сети.

В сети *INT* первое устройство является первым узлом, а последнее устройство перехода – конечным узлом. Все устройства, кроме первого и конечного узлов, являются промежуточными узлами. Каждое устройство *INT* играет разные роли в домене. Роли и их функции заключаются в следующем:

- Первый узел

Первый узел применяет политику уровня сервиса обслуживания (*Quality of service, QoS*) к входному порту трафика для зеркального отображения сообщения, соответствующего правилам, полученным на интерфейсе, на процессор *INT* в оборудовании. Процессор добавляет метку *INT* и временную метку к сообщению, генерирует сообщение *INT* и отправляет его на промежуточный узел.

- Промежуточный узел

После получения сообщения входящий порт промежуточного узла автоматически идентифицирует сообщение *INT* в соответствии с меткой,

отправляет его в процессор в оборудовании, добавляет временную метку, а затем отправляет его на конечный узел.

- Конечный узел

После получения сообщения входящий порт конечного узла автоматически идентифицирует сообщение в соответствии с меткой *INT* и отправляет его в процессор в оборудовании для добавления временной метки, а затем инкапсулирует сообщение *INT* в соответствии с указанным *IP*-адресом, номером порта и номером виртуальной локальной компьютерной сети (*Virtual local area network, VLAN*) и отправляет его в анализатор.

Мониторинг работоспособности сети

Работоспособность сетевых устройств – это комплексный показатель измерения. Чтобы построить алгоритм ИИ для проверки работоспособности сети, необходимо провести комплексный корреляционный анализ из нескольких измерений в соответствии с весовой долей сети, пользователя и приложения, тем самым получив оценку работоспособности, которая может точно отражать рабочие состояния текущей сети, пользователя и приложения. Когда показатель работоспособности высок, рабочее состояние относительно хорошее. В противном случае она эксплуатируется в состоянии частичной непригодности или даже в нерабочем состоянии.

Работоспособность сети состоит из трех измерений:

- Системная плоскость.
- Плоскость данных.
- Плоскость управления или менеджмента.

На рис. 1 продемонстрирована эталонная модель протоколов цифровой широкополосной сети интегрированных услуг (*broadband integrated services digital network, BISDN*).

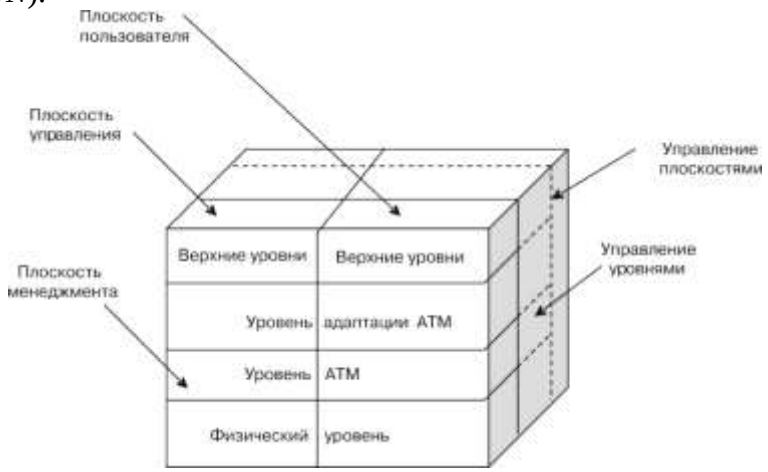

Рисунок 1

Показатели измеряются комплексно в соответствии с определенными весами, чтобы получить степень работоспособности сетевого устройства.

Работоспособность пользователя: мониторинг доступа пользователей к сети осуществляется с точки зрения следующих параметров:

- Беспроводная служба.
- Служба аутентификации.
- Сетевая служба.
- Службы управления сетью.

Таким образом, получается всеобъемлющий индекс работоспособности пользователя. Пользователи беспроводной и проводной связи оцениваются одинаково. Тенденция уровня работоспособности пользователя может показать качество входа в систему и качество подключения пользователей к сетевому ресурсу.

Работоспособность приложения: для получения индекса работоспособности приложения выполняется мониторинг трафика приложений и качества трафика. Индекс трафика в основном отражает доступность приложения, которая измеряется потерянными пакетами, индекс качества, в основном, отражает эффективность приложения, которая измеряется данными о задержке.

Многомерный корреляционный анализ: анализатор собирает различные типы данных, в основном включая состояние работы и передачи сетевых устройств, пользовательских терминалов, потоков приложений, а также конфигурацию служб и политик. Затем внутренняя взаимосвязь между данными интегрируется для формирования многомерных корреляций, которые помогают предоставлять своевременные предупреждения и обнаруживать проблемы и сбои в сети со следующих аспектов: анализ того, на какие серверы на стороне пользователя и на северные или южные потоки обслуживания повлияют аномалии.

Предупреждения для протоколов плоскости управления, таких как исключения подключения соседей по протоколу динамической маршрутизации (*Border gateway protocol, BGP*): анализ того, на какую службу между этими соседями повлияет исключение, и на доступ к службам серверов клиентов, которые будут затронуты.

Такие алгоритмы, как декомпозиция данных временных рядов и машинное обучение, могут поддерживать анализ аномалий и динамическое прогнозирование сетевых показателей. В частности, анализ аномалий проводится для обнаружения аномалий в прошлом, в то время как динамическое прогнозирование проводится для прогнозирования будущей тенденции данных и получения прогнозируемого графика. Для анализа аномалий используется динамический базовый подход для моделирования и обучения на исторических данных, собранных устройствами, чтобы сформировать динамическую базовую линию прогнозирования в качестве порога обнаружения аномалий. Базовая линия может быть улучшена за счет непрерывного машинного обучения и итеративной самокоррекции на основе исторических данных за определенный период. Алгоритм обнаружения аномалий, основанный на динамической базовой линии, может более точно отражать фактическую рабочую ситуацию в сети.

Динамическая базовая линия рассчитывается в автономном режиме через день на основе 14-дневных исторических данных, то есть прогнозируемое базовое значение на следующий день вычисляется за день до этого одновременно. Детализация сгенерированных динамических исходных данных соответствует детализации исходных данных.

Традиционное управление сетевым трафиком в основном остается на уровне состояния работы оборудования. Этот способ не может отражать реальное состояние пропускной способности сети и не может в полной мере показывать общую ситуацию с обслуживанием. Анализ сетевого трафика с использованием ИИ позволяет полностью воспринимать трафик ЦОД с глобальной точки зрения и формировать тесно связанную модель связи между сетью и бизнесом посредством сквозного визуального моделирования бизнеса, чтобы обеспечить всестороннюю техническую поддержку для оптимизации пропускной способности сети и раннего предупреждения о рисках сбоев работы оборудования.

С помощью технологии ИИ, машинного обучения и алгоритмов глубокого обучения, сеть становится более гибкой, анализатор производит исследование состояния сети с точки зрения приложения и активно распознает проблемы, существующие в сети и приложении. Он также обеспечивает возможность автоматического устранения неполадок в бизнес-задачах, быстро определяет границы и устраняет неисправности, а также снижает затраты на эксплуатацию и техническое обслуживание сети.

Стратегия настройки на основе ИИ относится к процессу анализа потоковых данных с зеркальным отображением *ERSPAN*, т.е. данных потока *INT* и показателей производительности телеметрии из измерений сети, приложения и пользователя, в соответствии с фактическим сценарием приложения сети ЦОД. Кроме того, алгоритмы ИИ, такие как динамическая базовая линия обнаружения аномалий, используются для выполнения интеллектуального анализа, который может активно определять, существует ли потенциальный риск наличия неполадок в сети, и обеспечивать раннее предупреждение. Сетевой анализ заключается в основном в проведении динамического мониторинга топологии сети, ее устройств и ресурсов устройств в режиме реального времени, оценке наличия ошибок с помощью интеллектуального анализа и раннего предупреждения. Примерами являются мониторинг трафика сетевого канала, мониторинг оптического модуля, мониторинг использования процессора и памяти устройства, мониторинг ресурсов чипа на уровне пересылки и т.д.

Анализ приложений фокусируется на определении того, является ли взаимодействие с идентифицированным приложением аномальным и является ли качество обслуживания приложения недостаточным, включая обнаружение аномалий *TCP*, визуализацию и анализ задержек путей пересылки приложений.

Анализ пользователей сосредоточен на пользовательском опыте, включая нарушения или сбои в доступе пользователей. Например, базовый анализ и обнаружение аномалий выполняются с использованием отношения числа пользователей, столкнувшихся со сбоями доступа, к общему числу связанных пользователей за определенный период времени в качестве ключевого показателя эффективности (*Key performance indicator, KPI*).

Заключение

Автоматическая оптимизация сетевой стратегии на основе ИИ позволяет инженерам всесторонне воспринимать условия сетевого трафика, время отклика на данные, статус передачи услуг и другую информацию посредством машинного обучения и анализа данных, а также автоматически настраивать сеть на основе изменений в сетевом трафике и состоянии работоспособности. Кроме того, автоматически выполняются проверки согласованности и целостности стратегии настройки, чтобы снизить вероятность ошибок и риски неполадок во время работы сети. Принципы работы анализатора, описанного в статье, может применяться в ЦОД на всех уровнях для повышения точности анализа трафика различных служб в центре обработки данных.

Литература

1. Смут С. Частные облачные вычисления: интеграция, виртуализация и сервис-ориентированная инфраструктура, Mechanical Industry Press, 2013.
2. Ванг К. Анализ протокола TCP/IP. Дж. Норт. Цзяотонский университет, 1 (1995), 112-117. DOI: CNKI: SUN: BFJT.0.1995-01-025.

3. Хе Я., Ли О., Ян Б.В., Лю Ю. Формальное описание протокола TCP на основе временной цветной сети Петри, Компьютер. Eng., 37 (2011), 77-80. DOI: CNKI: SUN: JSJC.0.2011-18-028.
4. Янг С., Гуи И.К. Стратегии анализа и управления сетевым трафиком, Вычисл. Ноу-хау. Технол., 10 (2011), 26-28+62. doi: CNKI: BC: BGDH.0.2011-10-012.
5. Болябкин М.В. Разработка и внедрение общего анализатора SQL // Международный журнал гуманитарных и естественных наук, 2022. – № 1-1(64). – С. 55-61. – DOI 10.24412/2500-1000-2022-1-1-55-61. – EDN PRDDVT.
6. Аббасов И.Б. Дизайн автономного мобильного робототехнического комплекса // Международный научно-исследовательский журнал, 2019. – № 1-1(79). – С. 33-40. – DOI 10.23670/IRJ.2019.79.1.006. – EDN YVLKWD.
7. Алексеев А.С. Методология стеганографической защиты с применением искусственных нейронных сетей // Вестник современных исследований, 2019. – № 1.13 (28). – С. 24-29. – EDN YWUDJR.
8. Алексеев А.С. Многоуровневая кибербезопасность интеллектуальных сетей // Вестник современных исследований, 2019. – № 1.13(28). – С. 30-34. – EDN YWUDJZ.
9. Андреева Е.А. Исследование математической модели оптимального управления динамической системой с использованием искусственной нейронной сети // Актуальные вопросы информатизации Федеральной службы исполнения наказаний на современном этапе развития уголовно-исполнительной системы: сборник материалов круглого стола, Тверь, 19 октября 2018 года / ФКУ «Научно-исследовательский институт информационных технологий ФСИН». – Тверь: Федеральное казенное учреждение «Научно-исследовательский институт информационных технологий Федеральной службы исполнения наказаний», 2018. – С. 343-347. – EDN YUFUVN.
10. Bessonova N.V. Artificial intelligence // Вестник Тульского филиала Финуниверситета, 2018. – No 1. – P. 364-365. – EDN RMNVNQ.
11. Васильев А.А. Термин «Искусственный интеллект» в российском праве: доктринальный анализ // Юрислингвистика, 2018. – № 7-8. – С. 35-44. – EDN YLQKSD.
12. Воронкова А.В. Интеллектуализация экономики как современный этап развития техники и технологии // Вестник Полоцкого государственного университета. Серия D. Экономические и юридические науки, 2018. – № 6. – С. 103-107. – EDN UZFDRO.
13. Гибадуллин А.А. Методы реализации игрового искусственного интеллекта // Современное программирование: Материалы I Международной научно-практической конференции, Нижневартовск, 15-18 ноября 2018 года / Ответственный редактор Т.Б. Казиахмедов. – Нижневартовск: Нижневартовский государственный университет, 2018. – С. 110-112. – EDN YUGQEX.
14. Гольдреер М.М. Адаптивный контекстно-тематический машинный перевод // Информационное общество, 2018. – № 3. – С. 59-65. – EDN SKFVCX.
15. Скуратов А.К. Статистический анализ телекоммуникационных сетей на основе исследования информационных потоков, представленных в виде временных рядов // Вестник Самарского государственного аэрокосмического университета, 2006. – № 1.– С. 259-262.
16. Высочина О.С., Шматков С.И., Салман Амер Мухсин. Анализ систем мониторинга телекоммуникационных сетей // Радиоэлектроника, информатика, управление, 2010. – № 2. – С. 139-142.
17. Бигелоу Дж. Стивен «Администрирование сети на основе Microsoft Windows 2000». Учебный курс MCSE. – М.: «Русская редакция», 2000. – 512 с.

18. Бигелоу Дж. Стивен «Сети. Поиск неисправностей, поддержка и восстановление». Учебное пособие. – СПб.: «БХВ-Петербург», 2007. – 415 с.
19. Новиков Ю.А. «Локальные сети: архитектура, алгоритмы, проектирование». Учебное пособие для ВУЗов, – М.: «ЭКОМ», 2000. – 568 с.
20. Норенков И.П., Трудоношин В.А. «Телекоммуникационные технологии и сети». Учебное пособие для ВУЗов. – М.: «ЭКОМ», 1999. – 392 с.
21. Олифер В.Г., Олифер Н.А. «Компьютерные сети. Принципы, технологии, протоколы». Учебник для вузов. 2-е изд. – СПб.: «Питер-пресс», 2002. – 864 с.
22. Олифер В.Г., Олифер Н.А. «Новые технологии и оборудование IP-сетей». Учебник для вузов – СПб.: «БХВ. – Санкт-Петербург», 2000. – 512 с.
23. Поляк - Брагинский А.В. «Администрирование сети на примерах». Учебное пособие. – СПб. «БХВ-Петербург», 2005. – 306 с.
24. Столлингс В.У. «Компьютерные сети протоколы и технологии интернета». Учебное пособие. - СПб.: «БХВ-Петербург», 2007. – 786 с.
25. Уилсон Э.Ж. «Мониторинг и анализ сетей». Учебное пособие. – М.: «Лори» 2002. – 504 с.
26. Храмцов П.Б. «Администрирование сети и сервисов Internet». Учебник для вузов. 2-е изд. – СПб.: «Питер-пресс», 2008. – 752 с.