DOI 10.37882/2223–2966.2020.04.27

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ АНАЛИЗА ТРАФИКА С ИСПОЛЬЗОВАНИЕМ МЕТРИКИ ХАУСДОРФА

DEVELOPING A MATHEMATICAL MODEL FOR TRAFFIC ANALYSIS USING HAUSDORFF METRICS

L. Makshanova
A. Tonkhonoeva
T. Tsybikova

Summary. This paper describes algorithms and methods for monitoring actual network traffic of a Telecom operator for loss or unauthorized access using network switches. Such algorithms and methods make it possible to control even larger amounts of data and dynamic processes on any network, which ensures the security and protection of the network operator. Based on the developed algorithm, software products for traffic control were implemented in the production process of Rostelecom's BF. In addition, the paper presents traffic statistics for the entire branch, which clearly shows the principle of traffic analysis. The definition of actual traffic is given. Illegitimate use of communication channels has been identified, which is accompanied by the appearance of abnormal areas in the time series. This article calculates the average traffic value, and also calculates the Hausdorff metric, the description and essence of which are also given in this paper. The calculation is based on the main formulas presented in this paper. It is also worth noting that to apply the traffic calculation method, it is important to have information and data about the approximate network structure; the number of subscribers in all network nodes; the distribution of subscribers by different classes of service; the list of network services; and the description of services. In addition, the article provides statistics of traffic analysis in the context of switches, it is given in the form of a figure.

Keywords: communication operator, network traffic, high-speed algorithm, traffic monitoring and analysis method, Hausdorff metrics.

Макшанова Лариса Михайловна
К.т.н., доцент, ФГБОУ ВО «Бурятский государственный университет имени Доржи Банзарова», Улан-Удэ
lorimak@list.ru
Тонхоноева Антонида Антоновна
Доцент, ФГБОУ ВО «Бурятский государственный университет имени Доржи Банзарова», Улан-Удэ
ant_ton@mail.ru
Цыбикова Туяна Сандаликовна
К.п.н., доцент, ФГБОУ ВО «Бурятский государственный университет имени Доржи Банзарова», Улан-Удэ
cts2001@mail.ru

Аннотация. В работе описываются алгоритмы и методы контроля действительного трафика сети оператора связи на предмет его потери или пропуска несанкционированного характера с помощью коммутаторов сети. Такие алгоритмы и методы дают возможность контролировать ещё большие объёмы данных и процессы динамического характера на любой сети, что обеспечивает безопасность и защиту сети оператора связи. На базе разработанного алгоритма были реализованы программные продукты контроля трафика в производственный процесс БФ ПАО «Ростелеком». Кроме того, в работе представлены статистические данные трафика всего филиала, где наглядным образом показан принцип анализа трафика. Дано определение собственно трафика. Выявлено нелегитимное применение каналов связи, которое сопровождается возникновением аномальных участков во временном ряду. В данной статье дан расчет среднего значения трафика, а также проведен расчет по метрике Хаусдорфа, описание и сущность которой тоже приведены в работе. Расчет произведен по основным формулам, представленным в данной работе. Также стоит отметить, что для применения методики расчета трафика важно обладать информацией и данными о примерной структуре сети; числе абонентов во всех узлах сети; распределении абонентов по разным классам обслуживания; списке сетевых услуг; описании услуг. Кроме того, в статье предоставлена статистика анализа трафика в разрезе коммутаторов, она дана в виде рисунка.

Ключевые слова: оператор связи, трафик сети, быстродействующий алгоритм, метод мониторинга и анализа трафика, метрики Хаусдорфа.

Введение

Сопровождение технико-экономического характера сетей требует решения целого комплекса задач, в том числе анализ и контроль управления сетью, мониторинг и качественное размещение сетевой инфраструктуры, обеспечение информационной безопасности в сетях, предупреждение о мошенничестве, управление сервисами, планирование и совершенствование новых услуг, обеспечение высококачественного обслуживания, увеличение удовлетворенности и лояльности клиентов.

Решение данных задач лежит в области управления процессом эксплуатации сетей операторов связи. Это одна из особенно важных и трудных задач. Компании в данной области изучают систему эксплуатации также

и с точки зрения эффективности ведения бизнеса, именно поэтому такой проблеме всегда уделяется довольно большое внимание. Также факторами, которые усиливают интерес к данной области, являются оценка рисков компании, ответственность по контрактам за оказание услуг, увеличивающаяся конкуренция на рынке операторов и связи, повышение требований клиентов к услугам связи и возможностям операторов быстро реагировать на изменение потребностей клиентов.

Вопросам оптимального проектирования и эффективной эксплуатации сетей уделяется большое внимание как в зарубежных странах, так и в России. Методы оптимизации и качественной работы сетей отражены во множестве работ и исследований различных специалистов данной области.

Собственно эффективность систем взаимодействия сетей зависит непосредственным образом от эффективности работы каких-либо информационных технологий. Для качественного предоставления услуг и поддержания своей конкурентоспособности операторы должны эффективным образом применять существующие у них ресурсные средства.

Актуальность проблемы усиливается с развитием объёмов трафика, ограниченностью региональных сетевых ресурсов, с повышением расходов на расширение полосы пропускания и потерь от несанкционированного пропуска трафика.

Чтобы выявить несвойственные моменты трафика, необходимо изучить использование метрики Хаусдорфа. Чтобы контролировать объемные и динамические процессы, имеющие место в сетях операторов связи, стоит применять быстродействующие алгоритмы и методы мониторинга и анализа трафика. Представление трафика в образе временного ряда дает возможность применять инструментарий прикладного анализа данных.

Собственно трафик представляет собой объем данных или число сообщений, которые были переданы с помощью канала за конкретный промежуток времени. Иначе говоря, трафик является большим числом телефонных разговоров и попыток установления соединения, которые проходят непосредственно через коммуникационные оборудования и/или телефонную сеть. Кроме того, трафик включает отношение между попытками вызова оборудования, который чувствителен к трафику, и скоростью выполнения данных вызовов.

Постановка задачи

Чтобы автоматизировать процесс нахождения моментов нелегитимного применения ресурсных элементов операторов, стоит разработать математическую модель, качественные для анализа временные ряды, которые включают в себя интервенцию и прочие признаки, свойственные для моментов злоупотреблений, автоматизировать анализ ряда дополнительных признаков и, при необходимости, задействовать оператора для принятия окончательного решения о важности проведения дополнительных тестов или блокирования недобросовестных абонентов.

Описание эксперимента

Чтобы проводить контроль процессов объемного и динамического характера, имеющих место в сетях операторов связи, стоит применять быстродействующие алгоритмы и методы анализа и контроля трафика. Представление трафика в образе временного ряда дает возможность применять инструментарий прикладного анализа данных (ПАД).

Также целесообразным образом необходимо отдельно изучить и оценить ряды с разнообразными аргументами, параметрами или показателями. К примеру, при изучении трафика операторов связи ими могут являться ряды, которые показывают загруженность каналов связи в фиксированные моменты времени, активность отдельных абонентов во времени, число и продолжительность звонков, которые проходят непосредственно через определенную станцию, активность абонентов отдельного тарифного плана и т.д.

Получаемые тем самым временные ряды считаются классическими объектами ПАД и для их изучения удобно применять аддитивную математическую модель.

Нелегитимное применение каналов связи сопровождается также возникновением аномальных участков во временном ряду, который является одним из видов трафика оператора. Всплески или провалы, резкие переходы к более низкому или наоборот высокому уже установившемуся значению могут являться свидетелями случаев и моментов выявления участков трафика с аномальным поведением и выявлением источников, которые считаются причиной возникновения данных участков, может потребоваться оценка многих других дополнительных параметров, проведение каких-либо тестов и испытаний, мероприятий, направленных на профилактику и т.д. для дифференциации нелегитимного применения ресурсных средств операторов от анормального поведения абонентов. К примеру, оборудование мошенников, которое работает в виде шлюза, не будет реагировать на входящие звонки по номеру, сопровождаемому исходящим трафиком.

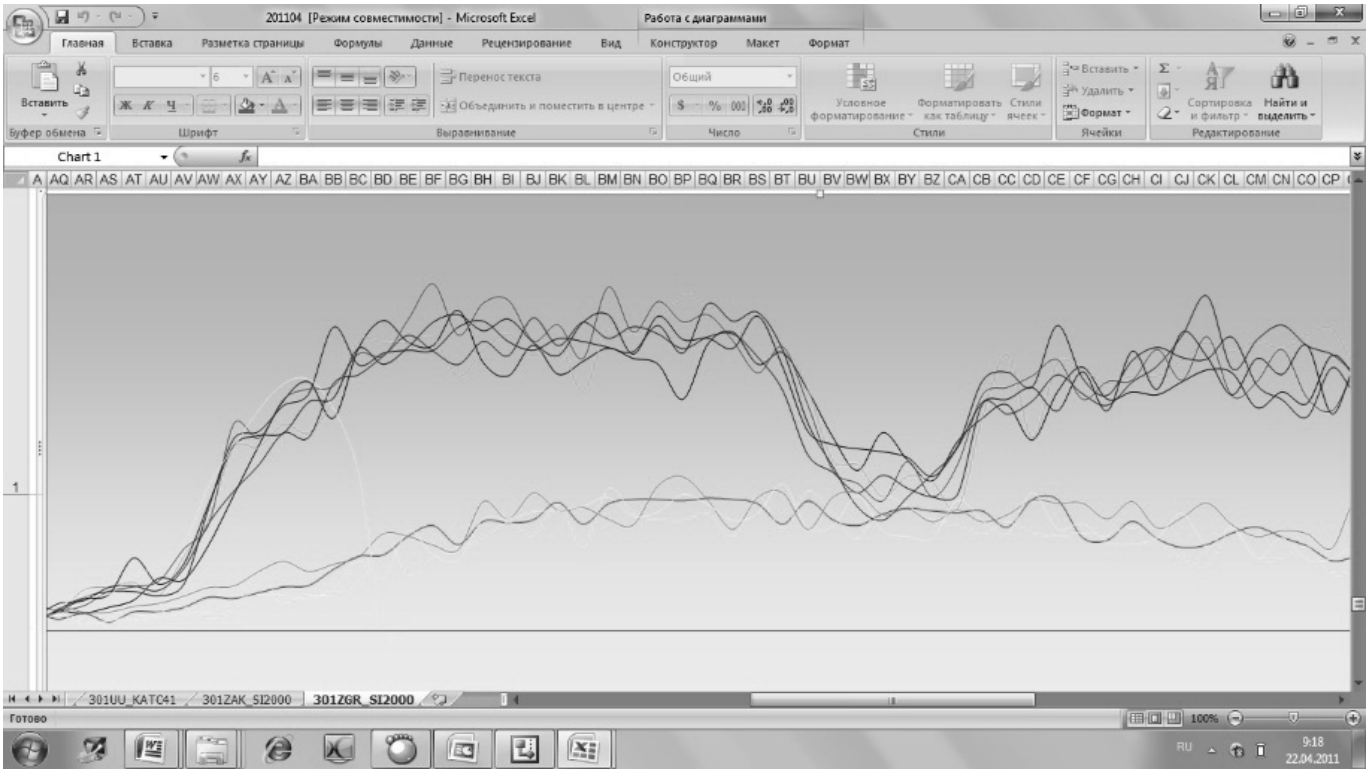

Рис. 1. Статистика анализа трафика в разрезе коммутаторов

Расчет среднего значения трафика

Основой расчета среднего значения трафика являются вероятностные свойственные характеристики потока данных, которые генерируются разнообразными сетевыми устройствами. Для применения такой методики важно обладать информацией и данными о:

- примерной структуре сети;
- числе абонентов во всех узлах сети;
- распределение абонентов по разным классам обслуживания;
- список сетевых услуг;
- описание услуг.

Трафик можно рассчитать отдельно для всех видов услуги на всех сетевых узлах. Для такого расчёта используется формула, которая представлена ниже следующим образом:

$$Y_i^{(k)} = B_{ср}^{(k)} * N_{аб\,i}^{(k)} * T_c^{(k)} * f_{вызов\,i}^{k}, \quad (3.1)$$

где, k — номер сетевой услуги; i — номер узла;

$Y_i^{(k)}$ — математическое ожидание трафика, который генерируется к-ой услугой на i-ом узле; $B_{ср}^{(k)}$ — скорость передачи данных (в битах на секунду);

$N_{аб.\,i}^{(k)}$ — количество абонентов на i-ом узле, которые используют k-ую услугу;

$T_c^{(k)}$ — средняя продолжительность сеанса связи для к-ой услуги;

$f_{вызов.\,i}^{(k)}$ — среднее количество вызовов в ЧНН для пользователей i-го узла, которые используют k-ую услугу.

Таким образом, среднее значение суммарного трафика можно будет рассчитать по следующей формуле:

$$Y_{ср} = \sum_{i=1}^{N}(\sum_{k=1}^{K} B_{ср}^{(k)} * N_{аб\,i}^{(k)} * T_c^{(k)} * f_{вызов\,i}^{k}), \quad (2)$$

где, N — количество узлов;

K — количество услуг (приложений).

Тогда доверительные границы трафика определяются по формуле:

$$Y_{дов} = Y_{ср} \pm D(Y), \quad (3)$$

где, $D(Y)$ — дисперсия трафика.

Кроме того, стоит разделить измерения за день по группам с одинаковым поведением статистического характера. В соответствии со спецификацией ITU-T данными группами считаются следующие показатели: рабочие дни, выходные и праздничные дни в году. Группирование

измерений с одинаковым статистическим поведением считается необходимым, поскольку дни с особенно высоким числом вызовов (к примеру, новый год или международный женский день) могут исказить итоговые данные.

Рекомендация отдела стандартизации телекоммуникаций E.492 [5] включает в себя также рекомендации по выявлению обычной и высокой интенсивности трафика в течение месяца. Согласно рекомендациям отдела стандартизации коммуникаций E.492 обычная интенсивность трафика в длительности месяца выявляется как четвёртый сверху наивысший пиковый трафик за день. Если же выбирать второй сверху наивысший результат измерений за месяц, что приводит непосредственно к увеличению интенсивности трафика за месяц. Данный результат дает возможность выявить прогнозируемую интенсивность трафика за месяц.

Метрика Хаусдорфа находит свое использование в задаче распознавания и сравнения величин как мера близости различных множеств и каких-либо объектов.

Для непосредственной оценки близости между компонентами, необходимо ввести данную метрику Хаусдорфа [6].

Расстояние Хаусдорфа между двумя графиками трафика f_1 и f_2 на отрезке $[t_1, t_2]$ можно определить непосредственно по следующей формуле:

$$p(f_1, f_2) = \max_{x \in [t_1, t_2]} |f_1(x) - f_2(x)|. \quad (5)$$

Такое расстояние, которое имеет название равномерной метрики, дает возможность увидеть, насколько значения одной функции изменились от другой.

При применении такой модели довольно легко разработать ПО анализа трафика на несвойственные всплески, если учитывать, что данный анализ применяет метрику Хаусдорфа.

Ниже на рисунке 1 представлена программа анализа с доверительными границами трафика для момента времени на отрезке $[t_1, t_2]$.

Заключение

Расстояние между двумя графиками трафика f_1 и f_2, которое рассчитывается непосредственным образом по метрике Хаусдорфа, имеет название равномерной метрики и показывает, насколько значения одной функции отклоняются от значения другой функции.

В том случае, если расстояние метрики на отрезке времени [t1, t2] будет превышать значение дисперсии трафика, то полученный в результате трафик будет идентифицироваться как несанкционированный доступ. Данный алгоритм процесса прогнозирования, анализа и контроля трафика дает возможность существенным образом снизить число потерь операторов за счет своевременного пресечения попыток несанкционированного применения его сетевых ресурсов.

ЛИТЕРАТУРА

1. Каграманзаде А. Г., Каграманзаде С. Д. Прогнозирование трафика — основа прогнозирования современных сетей электросвязи // ЦНТИ, "Информсвязь". № 1, М., 1991, 44 с.
2. Попков Д. Tranzit-fraud, или Мошенничество по крупному / ИнформКурьерСвязь. № 2. 2005.— с. 55–56.
3. Зарубин А., Седова Ю., Мошенничество на сетях связи / Connect! Мир связи, № 10. 2010. с. 106–109.
4. Hunter, Jane M. and Thiebaud, Maud E. Telecommunications Billing Systems: Implementing and Upgrading for Profitability (Professional Telecommunications) / N. Y.: McGraw-Hill –2003 г., 458 p.
5. Рекомендация ITU-T E.500. Принципы измерения интенсивности трафика. Рекомендация ITU-T E.492. Контрольный период для измерения трафика.
6. Хаусдорф Ф. Теория множеств.— 4-е изд.— М.: УРСС, 2007.— 304 с.
7. Макшанова Л. М., Бадмаева С. А., Анализ и оценка фрод-угроз сети оператора связи, Сборник научных статей по итогам международной научно-практической конференции, 14–15 августа 2015 года, г Санкт-Петербург: —Санкт-Петербург, 2015. С. 126–129.
8. Попков В. К. Сети связи и гиперсети // Методы и программы решения оптимизационных задач на графах и сетях.— Новосибирск, 1980. с. 77.

Журнал «Современная наука: актуальные проблемы теории и практики»