— цель — содержит краткое описание цели, на достижение которой направлена лабораторная работа;
— постановка задачи — содержит описание условий лабораторной работы и используемых элементов сети;
— схема сети — графическое изображение схемы сети, ее адресации, структуры и других необходимых для решения задачи элементов;
— краткие теоретические сведения — содержит минимально-необходимый теоретический материал, которым должен владеть обучаемый для успешного выполнения лабораторной работы;
— последовательность действий обучаемого — содержит описание основных действий (шагов) обучаемого, а также команды, которые необходимо выполнить для корректной настройки сетевого оборудования;
— дополнительное задание — содержит перечень необязательных задач, которые обучаемому предлагается решить самостоятельно.

Использование среды эмуляции Cisco Packet Tracer позволяет каждому студенту создавать, настраивать и исследовать собственную виртуальную вычислительную сеть, наблюдать основные процессы ее функционирования, а также наглядно изучать базовые сетевые протоколы.

ЛИТЕРАТУРА

1. *Cisco Systems, Inc.* Cisco SAFE reference guide [Электронный ресурс]. — `http://cisco.com/go/safe`.
2. *Convery S.* Network security architectures. Cisco Press, 2008. 792 p.
3. *Cisco Systems, Inc.* Cisco Packet Tracer [Электронный ресурс]. — `http://www.cisco.com/web/learning/netacad/course_catalog/PacketTracer.html`.

УДК 519.248

АДАПТИВНЫЙ АЛГОРИТМ ОТСЛЕЖИВАНИЯ АНОМАЛЬНОЙ АКТИВНОСТИ В КОМПЬЮТЕРНОЙ СЕТИ НА ОСНОВАНИИ ХАРАКТЕРНЫХ ИЗМЕНЕНИЙ ОЦЕНОК АЛЬТЕРНИРУЮЩЕГО ПОТОКА

О. В. Ниссенбаум, А. С. Присяжнюк

Современные сети связи, в том числе и компьютерные сети, представляют собой чрезвычайно сложные системы массового обслуживания, в которых функционируют различные потоки — заявок, пакетов данных, отказов и т. п. Как правило, такие потоки имеют переменную, случайно изменяемую интенсивность, поэтому могут быть моделированы простейшим потоком лишь на сравнительно небольших отрезках времени. В работе [1] показано, что трафик компьютерной сети с той или иной степенью достоверности может быть представлен дважды стохастическим потоком событий с кусочно-постоянной интенсивностью. В ряде случаев, например в распределенных информационных системах (ИС), где пользователи в основном работают с удаленными базами данных, трафик достаточно хорошо описывается моделью альтернирующего потока [2]. Такой поток имеет два состояния, в первом из которых наблюдается пуассоновский поток с параметром λ, а во втором события потока отсутствуют. Интервалы, на которых поток находится в первом или втором состоянии, распределены по экспоненциальному закону с параметром α_1 и α_2 соответственно.

Для компьютерных сетей, в том числе распределенных ИС, достаточно актуальными представляются исследования методов и алгоритмов отслеживания аномальной активности на канале, сегменте сети или локальной машине на основании анализа трафика. Выделяют два основных подхода в этой области: семантический и статистический анализ трафика. Пока круг приемов и методов статистического анализа весьма ограничен в сравнении с семантическим. Одной из причин является то, что традиционно использующаяся модель простейшего потока недостаточно хорошо описывает трафик реальной сети, а поэтому математические методы исследования, основанные на модели простейшего потока, недостаточно действенны.

1. Постановка задачи. Цель исследования — создание адаптивного статистического алгоритма, производящего мониторинг трафика на участке (отдельной машине, канале, сегменте сети и т. д.) и сигнализирующего о нетипичной для данного участка активности в режиме реального времени. Подчеркнем, что адаптивный статистический алгоритм предполагается использовать не как самостоятельное средство, а как дополнение к существующим методам для их прицельного применения.

В основу алгоритма положены следующие соображения. Трафик компьютерной сети достаточно хорошо приближается дважды стохастическим потоком событий, в частности альтернирующим потоком. Последний характеризуется тремя параметрами — λ, α_1 и α_2, которые возможно оценивать в режиме реального времени [3]. Характер трафика изменяется в зависимости от действий, производимых пользователями, а значит, изменяются и оценки параметров λ, α_1 и α_2. Отслеживая такие изменения, можно сделать вывод о типичности или нетипичности трафика на данном участке сети.

2. Сбор статистики трафика. Проведены наблюдения за трафиком компьютерной сети при различных видах активности пользователей. Фиксировались моменты поступления пакетов данных, на основании наблюдений каждые 10 с производились оценки $\hat{\lambda}$, $\hat{\alpha}_1$ и $\hat{\alpha}_2$. На рис. 1 в качестве примера приведена динамика оценок альтернирующего потока, полученных при наблюдении за входящим трафиком отдельного компьютера при работе с файлами в распределенной ИС в течение 40 мин.

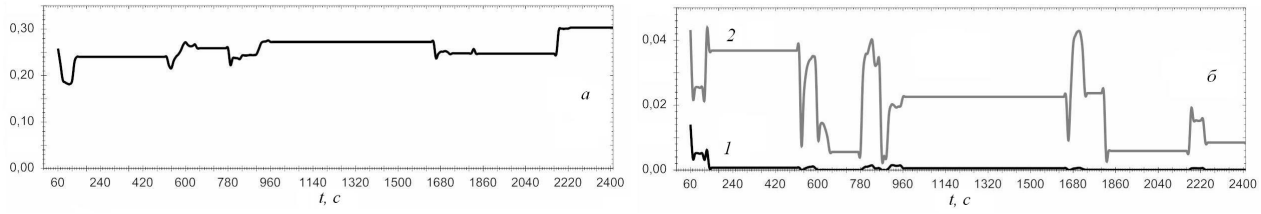

Рис. 1. Динамика оценок параметров входящего трафика компьютера при работе в распределенной ИС: *а* — $\hat{\lambda}$, *б* — $\hat{\alpha}_1$ (кр. *1*) и $\hat{\alpha}_2$ (кр. *2*)

3. Подходы к построению алгоритма. Для того чтобы сигнализировать о нетипичном поведении для трафика на сегменте сети, следует определить типичное поведение на этом сегменте. То есть необходим предварительный период наблюдений за трафиком. Таким образом, алгоритм должен работать в двух режимах:

1. Накопление данных, когда производится наблюдение за трафиком, оценка параметров и сбор информации о типичных значениях этих параметров. Предполагается, что в этот период трафик является «нормальным».

2. Отслеживание аномальностей, когда наряду с оценкой параметров и сбором информации производится сравнение текущих значений с прошлыми показателями и вывод о нормальности трафика.

При построении заявленного алгоритма следует учесть несколько важных моментов. Во-первых, это объем информации о трафике, который нужно хранить для функционирования алгоритма. Во-вторых, это изменение характера трафика с течением времени. В архитектуре и контенте сети могут происходить изменения, подчас значительные, такие, как включение в базы ИС новых полей, добавление или демонтаж серверов, перенос значительной части информации с одного сервера на другой. Такие изменения непременно скажутся на трафике, и ранее типичный трафик перестанет быть таковым. Наряду со значительными изменениями, могут происходить и небольшие, которые, накапливаясь, изменяют характер трафика постепенно. Это приводит к необходимости, во-первых, эффективного метода переключения алгоритма из режима накопления данных в режим отслеживания аномальностей и обратно, а во-вторых, постепенного вытеснения старых данных новыми.

Для анализа изменения оценок предполагается использовать следующие соображения. В каждый момент времени трафик описывается тремя числовыми значениями — оценками $\hat{\lambda}$, $\hat{\alpha}_1$ и $\hat{\alpha}_2$. Эту тройку можно представить точкой в трехмерном пространстве. Если характер трафика не претерпел существенного изменения на смежном интервале времени, то значения $\hat{\lambda}$, $\hat{\alpha}_1$ и $\hat{\alpha}_2$ также не должны существенно измениться, и тогда имеет место точка, достаточно близкая к предыдущей. Если же характер трафика резко изменится, то следующая точка будет значительно удалена от предыдущей. Производя наблюдения достаточно долго, получим несколько областей, в которых группируются точки (рис. 2). Если полагать, что весь период наблюдения трафик функционировал в нормальном режиме, то такие группы точек представляют собой эталон нормального поведения трафика для данного сегмента. Если же начиная с некоторого момента будут получаться точки, достаточно удаленные от ранее полученных, то это может служить сигналом об аномальной активности либо о том, что характер нормального трафика изменился. Для обработки данных, группирующихся по областям, можно использовать методы кластерного анализа [4].

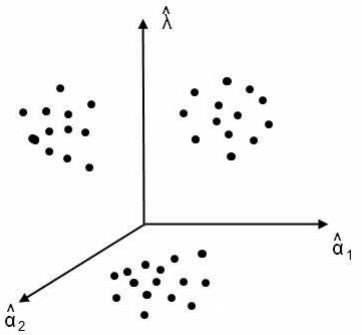

Рис. 2. Схема распределения оценок в координатах ($\hat{\lambda}$, $\hat{\alpha}_1$, $\hat{\alpha}_2$)

Выбросы траекторий оценок, наблюдающиеся, например, на рис. 1, могут затруднить кластеризацию. Причиной таких выбросов может служить как погрешность метода оценивания, отмеченная в [3], так и кратковременные изменения характера трафика. Например, на интервалах, когда трафик находится в переходном состоянии меж-

ду двумя режимами, на траекториях оценок возникают всплески. С одной стороны, следует отфильтровать такие выбросы, с другой стороны, их большое количество на интервале времени может сигнализировать об аномальностях.

Авторы предполагают использовать комбинированный подход, сочетающий статистический анализ, нейронные сети, методы теории графов, вейвлет-анализ и др.

4. Фильтрация оценок. Так как в ходе анализа траекторий оценок предполагается опираться на их высоко- и низкочастотную составляющие, то в качестве предварительной обработки данных следует разделить их. Для этой цели подходит разложение траекторий по базису Хаара [5] с последующим разделением высоко- и низкочастотной составляющих. На рис. 3 в качестве примера приведены низко- (сглаженные данные) и высокочастотная (выбросы) составляющие траектории $\hat{\lambda}$, представленной на рис. 1.

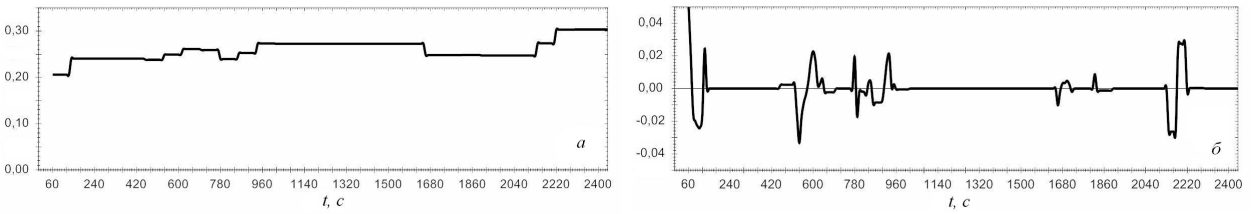

Рис. 3. Низко- ($а$) и высокочастотная ($б$) составляющие траектории $\hat{\lambda}$

Результаты и выводы. Сформулирована задача построения адаптивного статистического алгоритма отслеживания аномальной активности в компьютерной сети с использованием дважды стохастического потока в качестве модели трафика. Для случая распределенной ИС выбран альтернирующий поток, получено общее описание алгоритма, поставлены задачи, требующие решения в ходе построения алгоритма. Применен вейвлет-анализ для предварительной обработки данных.

Помимо решения задач, поставленных в рамках данной работы, открытыми остаются следующие вопросы: 1) выбор математической модели трафика в общем случае, когда альтернирующий поток недостаточно точно описывает его; 2) описание подходов к построению алгоритма в случае модели трафика, отличной от альтернирующего потока. Рассмотрение этих вопросов необходимо для построения адаптивного статистического алгоритма отслеживания аномальной активности для произвольной компьютерной сети, а не только для сети распределенной ИС.

ЛИТЕРАТУРА

1. *Головко Н. И., Каретник В. О., Танин В. Е., Сафонюк И. И.* Исследование моделей систем массового обслуживания в информационных сетях // Сиб. жур. индустр. матем. 2008. Т. XI. № 2(34). С. 50–58.
2. *Ниссенбаум О. В., Пахомов И.Б́.* Аппроксимация сетевого трафика моделью альтернирующего потока событий // Прикладная дискретная математика. 2009. Приложение № 1. С. 78–79.
3. *Васильева Л. А., Горцев А. М.* Оценивание параметров дважды стохастического потока событий в условиях его неполной наблюдаемости // Автоматика и телемеханика. 2002. № 3. С. 179–184.
4. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Verlag, 2009.
5. *Добеши И.* Десять лекций по вейвлетам. М.: Регулярная и хаотическая динамика, 2001.